Elucidating the molecular mechanisms conferred by the histone chaperone FACT

Inaugural-Dissertation

Zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Constantine Mylonas

aus Athen

Published: April 2018, Cologne

Gutachter: Prof. Dr. Andreas Beyer

Dr Peter Tessarz

Defence date: 15th November, 2017

Contents

1 Summary

1.1 English Version

The conserved and essential histone chaperone FACT (Facilitates Chromatin Transcription) reorganises nucleosomes during DNA transcription, replication and repair and ensures both, efficient elongation of polymerases and nucleosome integrity. In mammalian cells, FACT is a heterodimer, consisting of SSRP1 and SUPT16. Although the genetics and biochemistry of FACT are relatively well understood, it is not known whether it confers cell-type dedicated functions. However, genome-wide expression analyses across cell and tissue types implicate a role of FACT in maintaining an undifferentiated state. Here, we show that in mouse embryonic stem cells (mESCs), depletion of FACT leads to up-regulation and alternative splicing of pro-proliferative genes and key pluripotency factors concomitantly with hyper-proliferation of mES cells. Additionally, genes involved in neurogenesis are de-repressed in the absence of FACT, leading to faster progression of neuronal differentiation. To understand the role of FACT in regulating transcription at these loci, we performed MNase digestion of chromatin coupled to deep sequencing (MNase-seq) and Nascent Elongating Transcript Sequencing (NET-seq). Genes up-regulated upon FACT depletion, are characterised by loss of MNase-resistance just upstream of the transcription start site, suggesting that nucleosome deposition by FACT dampens gene expression. Finally, in combination with the NET-seq data, we show that this FACT-dependent nucleosome distribution at the promoter region is closely associated with strong uni-directionality of RNA Polymerase II towards the coding region. Taken together, FACT promotes Pol II governing through nucleosome deposition and thus enables maintenance of an undifferentiated state.

1.2 German Version

Das konservierte und essentielle Histon-Chaperon FACT (Facilitates Chromatin Transcription) reorganisiert Nukleosomen während der DNA-Transkription, Replikation und Reparatur und sorgt sowohl für eine effiziente Bewegung der Polymerasen durch das Chromatin als auch für die Nukleosomenintegrität. In Säugetierzellen formt FACT ein Heterodimer, bestehend aus SSRP1 und SUPT16. Obwohl die Genetik und Biochemie von FACT relativ gut verstanden sind, ist es nicht bekannt, ob es zelltypische und spezifische Funktionen verleiht. Allerdings implizieren genomweite Expressionsanalysen über Zell- und Gewebetypen eine Rolle von FACT bei der Aufrechterhaltung eines undifferenzierten Zustands. Hier zeigen wir, dass in murinen embryonalen Stammzellen (mESCs) die Depletion von FACT zu einer Hochregulierung und einem alternativen Spleißen von proproliferativen Genen und Genen kommt, die entscheidend für die Pluripotenz sind. Gleichzeitig sind viele Gene auch mit der Hyperproliferation von mES-Zellen assoziiert. Weiterhin sind andere Genen, die an der Neurogenese beteiligt sind, in Abwesenheit von FACT hochreguliert, was zu einer schnelleren Progression der neuronalen Differenzierung führt. Um die Rolle von FACT bei der Regulierung der Transkription an diesen Loci zu verstehen, haben wir MNase-Verdauung von Chromatin, gekoppelt an Micrococcal Nuclease-sequenzierung (MNase-seq) und *‚Nascent Elongating Transcript-Sequencing'* (NET-seq), durchgeführt. Bei der zweiten Methode wird die Position der RNA Polymerase nukleotidgenau und genomweit wiedergegeben. Genen, die nach der Depletion von FACT hochreguliert sind, zeichnen sich durch einen Verlust der MNase-Resistenz 5' der Transkriptionsstartstelle aus, was darauf hindeutet, dass die Nukleosomenablagerung durch FACT an dieser Stelle die Genexpression dämpft. Schließlich zeigen wir in Kombination mit den NET-seq-Daten, dass diese FACT-abhängige Nukleosomenverteilung an der Promotorregion eng mit einer starken Unidirektionalität der RNA-Polymerase II in Richtung der kodierenden Region assoziiert ist. Zusammengenommen läßt sich sagen, dass FACT die Aufrechterhaltung eines undifferenzierten Zustands durch die Positionierung von Nukleosomen und Regulation von RNA Pol II beeinflusst.

2 Introduction

2.1 Chromatin and nucleosomes

Chromatin is defined as a tightly packed and regulated nucleo-protein (DNA and histone proteins) complex that is readily accessible and can be found in the nucleus of eukaryotic cells. The basic functional repeating unit of chromatin is the nucleosome which is composed of ~147 base pairs (bp) of DNA wrapped around a histone octamer comprising of two units of each of the H2A, H2B, H3, and H4 histones (Schematic 1). All histones are subjected to different posttranslational modifications such as acetylation, phosphorylation, methylation, ubiquitination, and ADP-ribosylation. These modifications affect the biophysical properties of chromatin and have been shown to either promote or inhibit the functions of other factors. In addition, chromatin structure and especially nucleosomal properties can be robustly altered via the replacement of canonical core histones with alternate histone variants (Talbert & Henikoff, 2016), establishment or eviction of histones from DNA by ATP-dependent nucleosome remodellers (de Dieuleveult et al., 2016), and covalent histone modifications (Zentner & Henikoff, 2013).

Key biological processes in eukaryotic systems, such as development and differentiation, are constantly regulated by DNA-binding proteins that dictate transcriptional programmes in a cell-specific manner. Despite chromatin undergoing many transitions at many levels of biological organisation, three classes are identified, with regards to the chromatin aspects that affect epigenetic modifications;

i) specific histone marks associating with either promoter or enhancer regions,

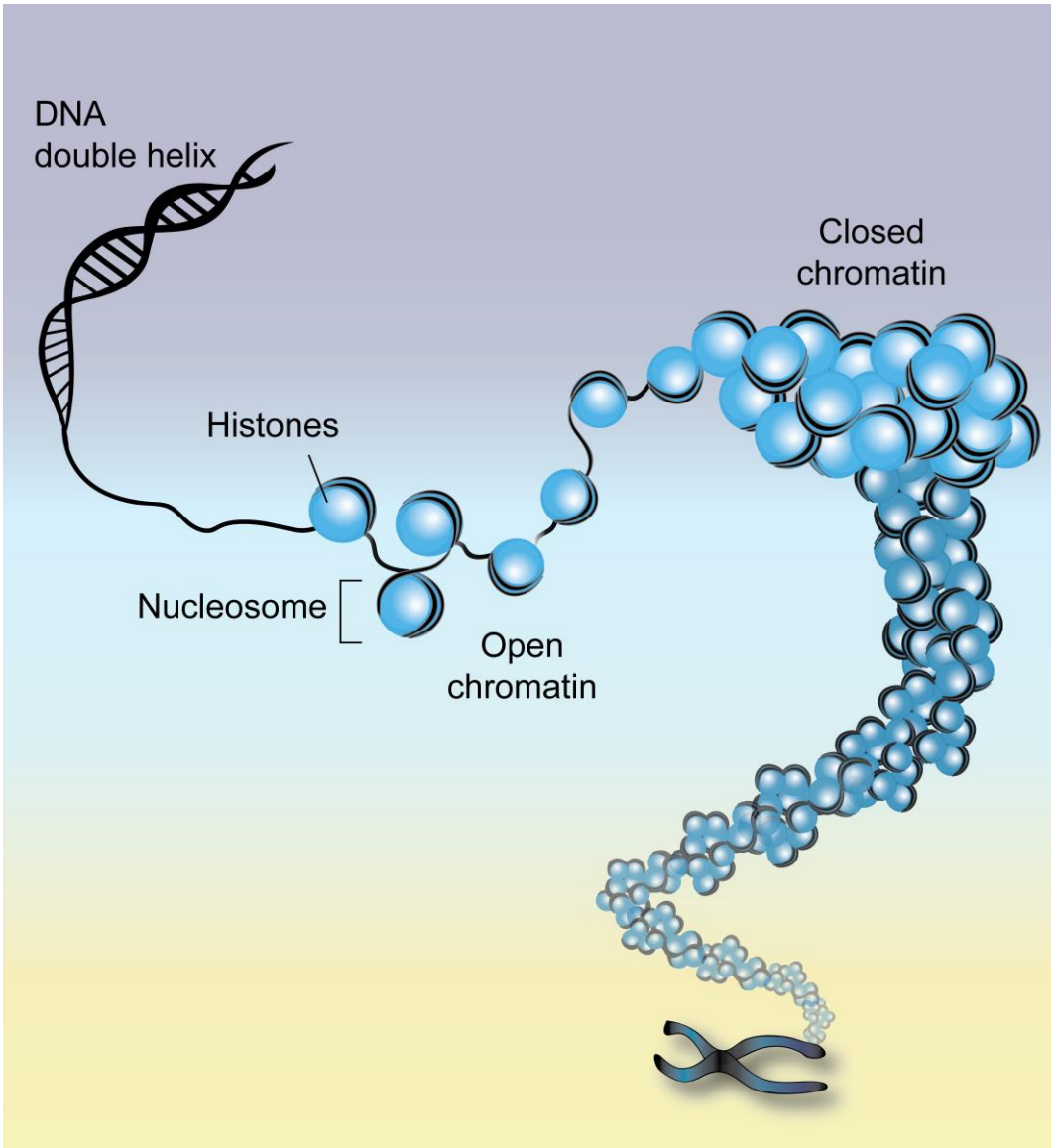
ii) selective localised access to regulatory elements (“open chromatin”) emerges as chromatin structures hinder entry to underlying DNA sequence motifs (Thurman et al., 2012) and

iii) long-range interactions between enhancers and targets regulate the nuclear architecture (Sanyal et al, 2012).

The advent of genome-wide chromatin immunoprecipitation-based techniques (ChIP-chip or ChIP-seq) has enabled the mapping of a vast amount of global histone modifications and transcription factors in many model organisms. A particular interest has arisen on modifications that are enriched over *cis*-regulatory elements. Trimethylation of H3K4 (H3K4me3) and H3K36 (H3K36me3) is a mark highly enriched on the promoters and gene body of actively transcribed genes, respectively.

Other marks, such as H3K27Ac and H3K4me1 are usually localised over putative enhancers. Impairment of methyltransferase complexes leads to drastic reductions in H3K4me3 and H3K36me3 levels but the impact on the transcriptional landscape is minimal, suggesting that transcriptional regulation is not directly regulated by histone marks (Zentner & Henikoff, 2013).

In fact, it has been suggested that posttranslational modifications may function as regulatory modules to some extent. Indeed, establishment of H3K36me3 on the gene body area has been linked to nucleosome stability and repression of cryptic transcription (Keogh et al., 2005). In general, histone modifications augment changes in nucleosome occupancy mediated by transcriptional elongation and chromatin remodelling factors.



Schematic 1: Chromatin compaction model. The DNA double helix is wrapped around histones forming the nucleosome; the basic structure of chromatin. Depending on the sparsity of nucleosomes across the genome, chromatin can be characterised as open (euchromatin) or closed (heterochromatin). This DNA-protein structure further undergoes further compaction and forms chromosomes that fit into the nucleus and hold the all the genetic information of an organism.

2.2 Chromatin structure and dynamics

The presence of histone marks also defines the status of chromatin; either highly active and permissive (euchromatin) or closed and repressive (heterochromatin). Despite the non-permissive nature of heterochromatin, structural repressive DNA domains vary from one cell type to another, hence defining this type of heterochromatin as facultative (H3K27me3). In contrast, constitutive heterochromatin (H3K9me3, and H4K20me3) is permanently non-accessible among different cell types and usually devoid of genes as it is established around the centromeric and telomeric regions of chromosomes.

Over the past 20 years, genetic screens in flies, yeast, and plants have identified key factors responsible for chromatin-dependent gene regulation, such as heterochromatin protein 1 (HP1), Suppressor of variegation 3 –9 (Su(var)3–9), Polycomb (Allis & Jenuwein, 2016). Alas, the molecular function of these factors with regard to enabling a euchromatin to heterochromatin transition, and vice-versa, remains elusive.

A successful interaction of transcription factors (TFs) with DNA regulatory elements requires the reorganisation of nucleosomal structures. Such disturbances in the nucleosomal landscape can be easily mapped via the detection of DNA sequences characterised by increased nuclease accessibility (DNase hypersensitive sites - Thurman et al., 2012). These “chromatin-permissive” events are dedicated to cell-specific functions and govern diverse transcriptional programmes.

Apart from the “canonical” histones that comprise the core unit of the nucleosome, evolutionary mechanisms have led to the emergence of histone

variants that bestow chromatin with special properties in a locus-specific manner. Core and variant histones display major differences with regard to spatial organisation and temporal expression pattern along the linear genome, with the histone variant-containing nucleosomes exhibiting specific and unique distributions. Histone deposition and eviction are necessary in modulating chromatin organisation and ultimately, DNA template-dependent biological functions. Hence, the evolution of specific chaperones and ATP-dependent remodellers has been a critical component in catalysing the above changes in chromatin structure (Buschbeck & Hake, 2017). All chromatin remodellers share a conserved helicase-like ATPase motor that powers all subsequent remodelling actions. However, they lack strong DNA-specific binding motifs as opposed to TFs. A chromatin remodeller – DNA contact is mediated via nucleosome binding through their N-terminal domains.

Given the nucleosome-specific function of remodellers, it is not surprising that functional interactions also exist between chromatin remodellers and histone modifications. Deletion of histone acetyltransferases (HAT) is synthetically lethal with loss of RSC (Remodels the structure of chromatin) or SWI/ SNF (SWItch/ Sucrose Non-Fermentable) (Kasten et al., 2004). It is tempting to speculate a synergistic effect between remodellers and histone modifications.

2.3 Chromatin dynamics in early embryogenesis

Embryonic cells acquire a cellular identity right after fertilisation that enables them to commit to a specific lineage. Yet, it is unknown how chromatin conformation within each cell promotes this commitment throughout the whole developmental cycle. Nevertheless, it is postulated that embryonic chromatin structure regulates cell fate by operating at two levels via the regulation of genes coding for lineage specific TFs and the alteration of chromatin dynamics and nuclear organization on a global scale (Burton & Torres-Padilla, 2014).

A *de novo* chromatin assembly right after fertilisation results in the incorporation of newly synthesised histones (mainly hyperacetylated and hypomethylated) deriving from the maternal genome (Burton & Torres-Padilla, 2014). Thus, the newly assembled genome is devoid of heterochromatin. This loss of heterochromatin is consistent with a more open chromatin conformation characterising pre-implantation development (Ahmed et al., 2010) and the maintenance of a pluripotent state. In general, it has been previously proposed that this “open” chromatin conformation allows transcriptional programmes to swiftly switch upon the induction of a differentiated cell state (Gaspar-Maia et al., 2011).

2.4 Chromatin and pluripotency

The idea of an open chromatin in ES cells has been initially supported by electron microscopy, where heterochromatin levels in differentiated cells exhibited significantly higher levels of compact chromatin compared to ES cells (Park et al., 2004). In addition, ChIP-seq analyses of both constitutive (H3K9me3) and facultative (H3K27me3) heterochromatin marks further support the prevalence of a more “compact” chromatin in a differentiated state (Hawkins et al., 2010).

The presence of active gene expression marks (H3K4me3, H3K27Ac) is more prominent in ES cells as opposed to somatic cells. However, there are counter-acting mechanisms that repress developmental regulatory genes, and hence maintain a pluripotent state. These developmental regulators are usually poised, i.e. they are characterised by the presence of both active (H3K4me3), and repressive (H3K27me3) histone marks (Azuara et al., 2006; Pan et al., 2007). This bivalent state allows, during differentiation, the rapid expression of lineage-specific genes through loss of the repressive H3K27me3 mark, or the repression of genes through the loss of H3K4me3.

Moreover, apart from active histone marks, open chromatin in ES may be maintained by ATP-remodellers and histone chaperones via nucleosome assembly/ disassembly as well as the unwinding of highly ordered structured chromatin. Taken into consideration the high expression levels of remodellers in ES cells (de Dieuleveult et al., 2016; Efroni et al., 2008) along with the abundance of active histone marks, it is possible that this collaboration orchestrates an open chromatin state on a global scale.

2.5 Histone chaperones

Core histones have the ability to bind DNA *in vitro*, thus resulting in the formation of a diverse array of chromatin structures. A proper co-ordination of histone dynamics with regard to the assembly of well-defined nucleosomes is being conferred by key factors called histone chaperones. Histone chaperones are involved in multiple steps of nucleosome formation and maintain histone stability or degradation (Burgess & Zhang, 2013). Potential functions involve;

1) regulating interaction of histones with importins,

2) direction of specific enzymes to histones for the establishment of specific post-translational marks (PTMs),

3) facilitation of intra-histonic interactions for nucleosomal formation and stability (Buschbeck & Hake, 2017). On the other hand, chromatin remodellers use ATP to catalyse changes in chromatin structure and can allow the exchange of core histones with histone variants (Narlikar, Sundaramoorthy, & Owen-Hughes, 2013).

Although histone chaperones share the common attribute of binding histones (either H3-H4 or H2A-H2B oligomers), a few of them can bind to specific histones (canonical or variants) alone (Obri et al., 2014). Despite this specificity, most histone chaperones play a crucial role in stabilising nucleosome oligomers in a free soluble state; a critical step for nucleosome formation. In addition, a collaboration between histone chaperones, and chromatin remodellers can influence gene expression by altering nucleosomal stability via canonical or variant histone deposition and thus enable/prevent the recruitment of TFs.

2.6 The histone chaperone FACT

Several histone chaperones involved in nucleosome assembly have been shown, over the past two decades, to fundamentally alter chromatin structure, thus allowing crucial DNA-based processes such as DNA repair, replication, and transcription to take place by safeguarding non-nucleosomal histone-DNA interactions. They mainly bind to regions on the histones that are crucial for nucleosome formation, thus indicating a putative role in the stabilisation of oligomers in the free and soluble state adjacent to the nucleosome. In contrast to ATP-dependent remodelling factors, which use energy from ATP hydrolysis to disrupt histone-DNA contacts, nucleosome destabilisation by histone chaperones is mediated through the use of the spontaneous movement of the DNA around the dyad axis (Hondele et al., 2013). Moreover, histone chaperones, in tandem with ATP-dependent chromatin remodellers, deconvolute the organisation of highly ordered chromatin in order to enhance accessibility of cellular components, such as Pol II, thereby facilitating transcription (Venkatesh & Workman, 2015).

The histone chaperone FACT (Facilitates Active Transcription) is a 180k Da heterodimer, comprising of SSRP1 (Structure Specific Recognition Protein 1) and SUPT16 (Suppressor of Ty 16) proteins, that promotes polymerase progression via nucleosome destabilisation (Belotserkovskaya et al., 2003; Orphanides et al, 1999). Although the structure of FACT is highly conserved among eukaryotes, in yeast and fungi the SSRP1 subunit (Pob3 in yeast) lacks DNA binding affinity. Instead, DNA binding capacity is implemented by Nhp6 that loosely associates with the FACT complex and promotes its function *in vivo*; the displacement of H2A-H2B dimers to allow passage of the transcribing RNA polymerase (Winkler

& Luger, 2011). Interestingly, Nhp6 serves also as DNA binding mediator for other elongation factors, such as SPT6 (Suppressor of Ty 6).

Deletion of FACT in yeast leads to inviability (Jeronimo et al., 2015) whereas a homologous deletion of *Ssrp1* in mouse embryos compromises survival after preimplantation (Cao et al., 2003). Concomitantly, depletion of FACT in human cancer cells hinders cell proliferation indicating that FACT might serve as a potential drug target against tumorigenesis (Gasparian et al., 2011; Koman, Commane, & Paszkiewicz, 2012).

Generally, in higher eukaryotes, FACT is highly enriched in stem and cancer cells while its levels are significantly reduced in differentiated cells. This suggest a role of FACT in maintaining an undifferentiated state (Garcia et al., 2011).

2.7 Transcriptional regulation by FACT

FACT activity is crucial for transcriptional elongation and is enabled by two distinctive mechanistic models; (i) Histone removal in front of RNA Pol II, and (ii) nucleosome reassembly in its wake. The above processes have been elucidated *in vitro* and pinpoint;

a) A strong affinity of FACT with nucleosomes via direct interaction with H2A-H2B dimers that allows efficient transcriptional elongation (Orphanides et al., 1999).
b) The disassembly of H2A-H2B dimers by FACT (Belotserkovskaya et al., 2003).

Corroborative evidence supports the above notion. In yeast, mutations that weaken the interaction between H2A-H2B and (H3-H4) tetramers can suppress defects in FACT function both *in vivo* and *in vitro*, reinforcing the role of FACT in promoting histone eviction (McCullough et al., 2011). A major debate in the field still rises with regard to whether H2A-H2B deposition is a directed and necessary result of FACT activity (Duina, 2011). As a result, apart from the orginal “dimer eviction model”, a second “non-eviction/ global accessibility model” suggests that the FACT-nucleosome complex results to tethering of all histone subunits in a specific conformation state that allows easier accessibility and loss of H2A-H2B dimers (Formosa, 2012). In this latter model, the eviction of H2A-H2B is not a consequence of FACT activity but instead it is mediated by extrinsic factors such as the force exerted by the elongating RNA Pol II.

2.8 Chromatin and FACT

Previous reports pinpoint the involvement of FACT in transcriptional initiation and elongation through binding to the H3K4me3 mark, Chd1 chromatin remodeller, and Pol II (Sims et al., 2007). Moreover, FACT inactivation in yeast causes increased transcription through elevated Pol II and TATA-binding protein (TBP) occupancy (Mason & Struhl, 2003). This suggests a putative function of FACT in suppressing cryptic transcription at certain promoters.

FACT is also influenced by the chromatin environment across transcribing genes by altering histone modifications. Mono-ubiquitination of H2B can mediate FACT-dependent histone eviction during transcriptional elongation. On the contrary, mono-ubiquitination of H2A prevents transcriptional elongation by inhibiting FACT recruitment to chromatin (Zhou et al., 2008). Recently, it has also been observed that glutamine methylation (H2AQ105me) restricts binding of FACT on ribosomal DNA (Fig 1 - Tessarz et al., 2014).

In addition to being a key contributor to transcriptional elongation, FACT is also involved in mRNA nuclear export (Hautbergue et al., 2009) as well as the efficient splicing of transcripts (Burckin et al., 2005).

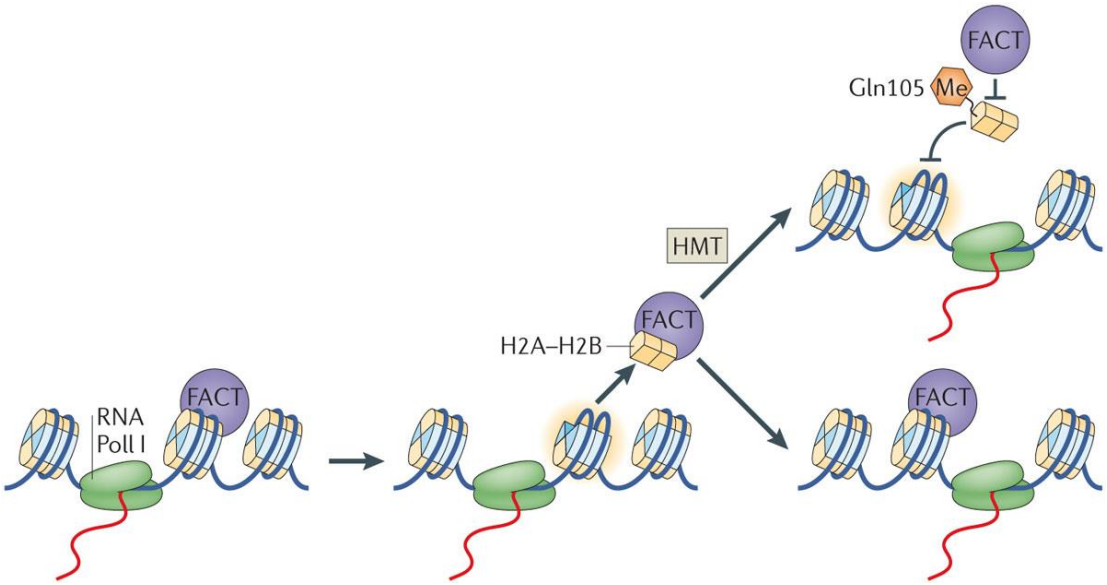


Figure 1: FACT facilitates passage of the polymerase by remodelling chromatin. Methylation (me) of Gln105 (Q105me) in H2A prevents FACT from interacting with the H2A-H2B dimer and thus leads to decreased nucleosome reassembly in the wake of RNA polymerase I (RNA Pol I). Consequently, the formation of a more open chromatin takes place, devoid of nucleosomes, which ensures high levels of rDNA transcription. HMT, histone methyl transferase. (Taken from Tessarz & Kouzarides, 2014).

2.9 Transcription mechanics by RNA polymerase II

RNA polymerase II (Pol II) is responsible for the transcription of DNA into mRNA with the aid of additional protein factors. Transcription is usually divided into three phases on the basis of the factors and the mechanisms that regulate Pol II. One of the most well studied mechanisms is the reversible phosphorylation of the largest subunit of Pol II; the carboxy-terminal domain (CTD) which contains a repeating heptapeptide sequence (YSPTSPS). The different factors and CTD-phosphorylated residues that associate with Pol II in a stage-specific manner are summarised in Table 1. The initiation phase involves binding of Pol II (S5 ph) to the gene promoter region (Sainsbury, Bernecky, & Cramer, 2015), followed by the elongation phase, during which RNA synthesis occurs. Nevertheless, higher eukaryotes display an additional signal integration step that separates the two aforementioned phases and keeps the polymerase paused at promoter-proximal regions (Venkatesh & Workman, 2015). Transcriptional elongation is followed by termination, where the transcribed RNA is released from Pol II (S2 ph). This cyclic and continuous repetition of all three phases determines a gene's expression levels and ensures normal cellular and organismal functions.

The completion of transcriptional initiation does not necessarily guarantee a robust transition into transcriptional elongation. Pausing of RNAPII can be observed after the transcription of 20 and 60 nucleotides in almost 30% of the total human genes. However, this pausing is temporary, and hence allows Pol II to resume transcriptional elongation (I Jonkers & Lis, 2015).

Pausing of Pol II at promoter-proximal regions is highly dependent on the core promoter features responsible for Pol II recruitment. This is accomplished via a synergistic effect between promoter-associated transcription factors (TFs), the

negative elongation factor (NELF), and DRB-sensitivity-inducing factor (DSIF) (Table 1). Phosphorylation of NELF/DSIF by the positive transcription elongation factor (P-TEFb) complex enables the release of paused Pol II and mediates the progression of transcriptional elongation. P-TEFb has also been shown to phosphorylate the elongating Pol II variant (Pol II S2), thus allowing promoter-proximal pause release (Iris Jonkers, Kwak, & Lis, 2014) (Schematic 2).

Although promoter-proximal pausing occurs in a wide range of eukaryotes, its function still remains to be elucidated. Jonkers and Lis (2015) propose three non-exclusive models for its function: (1) The presence of a stalled Pol II is a characteristic of an active and open promoter; (2) Changes in gene expression are highly dependent on stimuli response. Therefore, gene expression can be accelerated by skipping transcriptional initiation and Pol II can enter a paused state. As a result, genes that display a higher paused Pol II will be activated more swiftly; (3) Due to co-transcriptional processing, Pol II is associated with several other proteins during transcription. The levels of pausing depend on a balance between pausing and activating factors.

Transcriptional elongation is more complex than initially thought. Pol II elongation rates fluctuate between and within genes, and seem to facilitate splicing, termination, as well as genomic stability (Iris Jonkers et al., 2014). Transcriptional elongation by Pol II is a highly dynamic process and an indispensable part of the transcription cycle. ChIP of RNA pol II followed by sequencing, maps distribution of the polymerase across the genome and provides great insight with regard to RNA transcription. Nevertheless, caveats such as low resolution and vague strand specificity predominate. In addition, ChIP–seq of Pol II displays high

background levels on the gene body areas, and its use is restricted to highly expressed genes (I Jonkers & Lis, 2015). Hence, a number of techniques have been developed to study elongation rates on a genome-wide scale (Kwak et al., 2013; Mayer & Churchman, 2016; Min et al., 2011). In Global run-on-sequencing (GRO-seq) and Precision run-on-sequencing (PRO-seq), Pol II is halted and then restarted *in vitro* under the presence of modified nucleotides that enable nascent RNA purification. Transcription run-on based methods are highly dependent on the transcriptional restart of Pol II under non-physiological conditions; thus avoiding detection of Pol II that is in the process of pause-recovery.

Native elongating transcript sequencing (NET-seq) can capture nascent RNA transcripts at high resolution and detect unstable anti-sense transcripts before they are turned over by the cellular RNA degradation pathways (Mayer et al., 2015). This allows high resolution mapping of RNA polymerase (paused, recovering, and not paused) during initiation and elongation. It also enables the characterisation of distinctive regions of high Pol II pausing where splicing of RNA transcripts occurs.

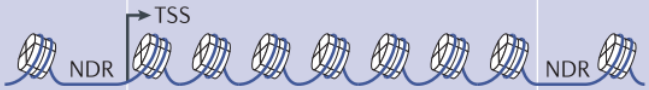
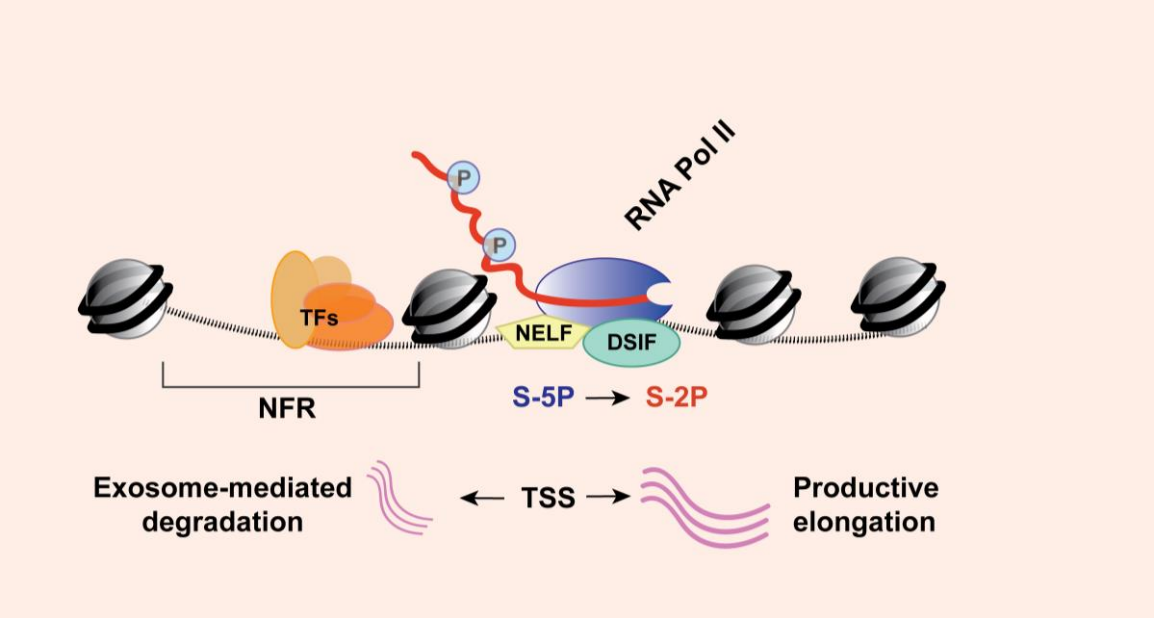
Stage of transcription	Initiation	Elongation	Termination
Pol II CTD phosphorylation status	S5P and S7P	S5P (5′ end of gene), Y1P and S2P(3′ end of gene)	S2P
Gene regions	Promoter	Coding region 	Terminator
Stage-specific regulatory factors	• Activators • Pol II • Mediator • Basal transcription factors • Termination factors • Chromatin modifiers • Chromatin remodellers • Histone chaperones	• Pol II • CTD kinases and phosphatases • Elongation factors: DSIF (Spt4 or Spt5), NELF and PAF complex • Chromatin modifiers • Chromatin remodellers • Histone chaperones	• Pol II • Termination factors • CTD phosphatases • Chromatin remodellers

Table 1: The CTD domain of Pol II serves as a protein-recruiting platform. Different phosphorylation statuses in the heptapeptide sequence (YSPTSPS) of the CTD domain of Pol II are crucial for transcriptional initiation, elongation, and termination. A diverse combination of kinases, histone chaperones, chromatin remodellers, and phosphorylated residues work in synergy for an efficient activity of Pol II (Venkatesh & Workman, 2015).



Schematic 2: Promoter architecture model. A typical eukaryotic promoter includes a nucleosome free region (NFR) upstream of the TSS where binding of most transcription factors (TFs) occurs. RNA Pol II is stalled by the NELF/DSIF. Its tail phosphorylation at the 5th Serine (S5 ph) indicates proximal promoter pausing. When NELF/DSIF dissociates, Pol II is phosphorylated at the S2, thus marking its progression towards transcriptional elongation. Pol II can also transcribe on the opposite direction. However, those transcripts are immediately degraded by the exosome complex.

2.10 Alternative splicing of RNA transcripts

Alternative splicing is defined as the process by which exons from a single gene can be either included or excluded from the final mature RNA (mRNA) product, making it an important regulatory crossroad between transcription and translation. Alternative splicing affects over 95% of all mammalian genes and multiple regulatory processes such as chromatin modifications and signal transduction (Kornblihtt et al., 2013).

The spliceosome is a multi-ribonucleoprotein complex that assembles at the splice sites of each intron and mediates the splicing process. A consensus sequence at each splice site is fundamental not only for recognition by the spliceosomal components but also for assessing the “strength” of the splicing site (Sibley, Blazquez, & Ule, 2016). The competition between “strong” and “weak” splice sites along the nascent pre-mRNA leads to alternative splicing of transcripts.

The 5’ and 3’ splice sites mark the beginning and the end of each intron and undergo “cutting and sewing”. A combination of small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5, and U6 and auxiliary factors (U2AF65, U2AF35) form the spliceosomal complex that performs two transesterification reactions necessary for intron excision and subsequent joining of selected exons (Fig 2). Together with RNA-dependent ATPases and helicases, the spliceosome transitions from an inactive to a catalytically active state.

Alternative splicing is more prevalent in multicellular eukaryotes suggesting its significant biological importance with regards to the vast expansion of coding capacity that it provides. This divided organisation of eukaryotic genes into exons and introns seems to have two advantages;

- Formation of new genes via non-disruptive recombination events that allowed the fusion of protein-coding exons from different ancestor genes. This mutational process is known as exon shuffling enables new genes to carry over the splicing signals of the ancestor genes (Hynes, 2012).

- Expansion of coding capacity of eukaryotic genomes. A single gene can produce two or more mRNAs that are very similar but not identical. Subsequently, this allows the translation of different proteins.

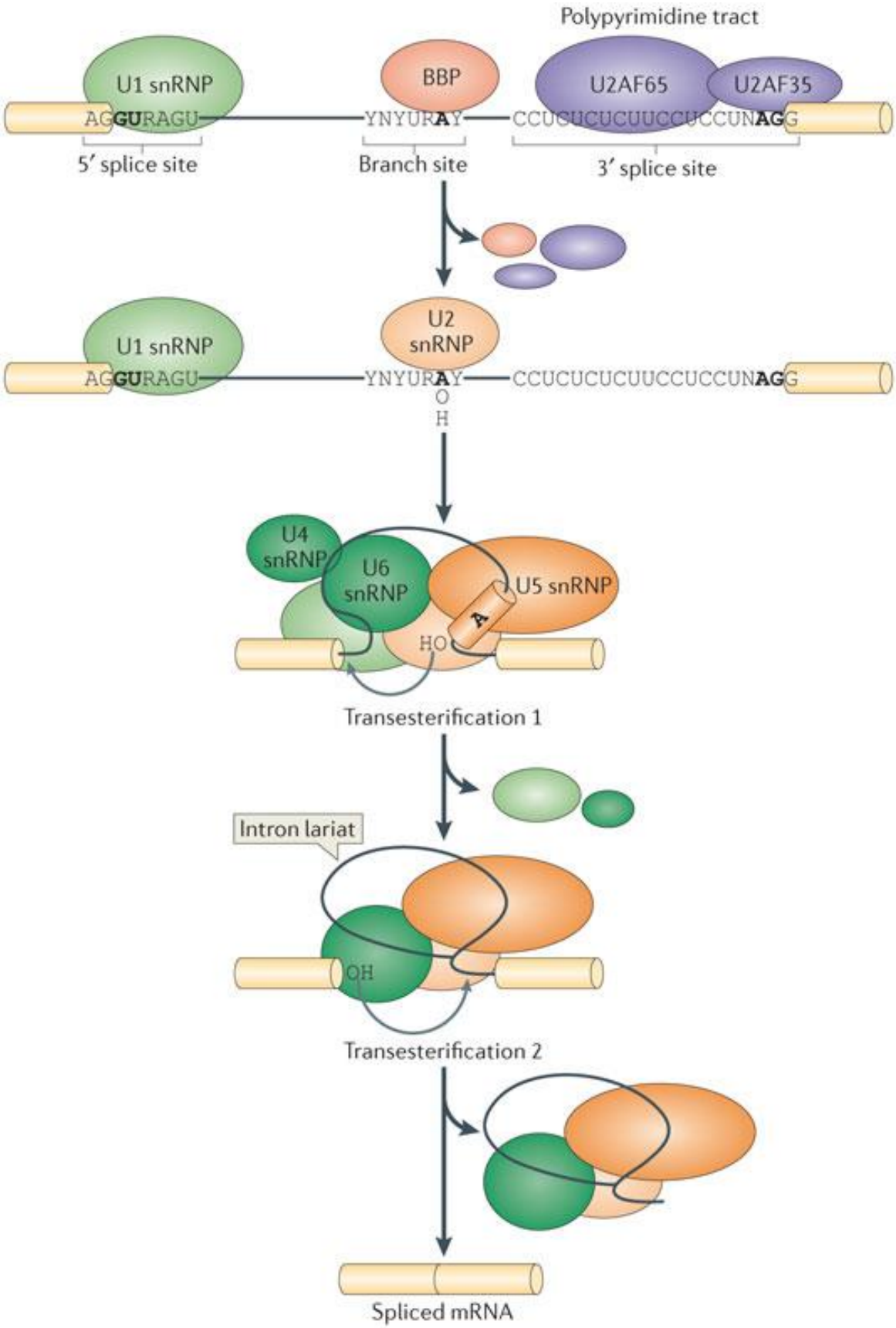

Polypyrimidine tract
U1 snRNP
BBP
U2AF65
U2AF35
AGGURAGU
YNYURAY
CCUCUCUCUUCCUCCUNAGG
5′ splice site
Branch site
3′ splice site
U1 snRNP
U2 snRNP
AGGURAGU
YNYURAY
CCUCUCUCUUCCUCCUNAGG
O
H
U4 snRNP
U6 snRNP
U5 snRNP
A
HO
Transesterification 1
Intron lariat
OH
Transesterification 2
Spliced mRNA

Figure 2: The spliceosome mediates a two-step splicing reaction. Initially, the branch site, the 5' splice site, and the 3' splice site are recognised by the branchpoint binding protein (BBP), the U1 snRNP, and the auxiliary factors (U2AF65, U2AF35), respectively. The first transesterification step involves a nucleophilic attack by the 2' OH group of a key adenosine in the branch consensus site on the 5´ splice site. This results to the formation of a branched RNA intermediate known as the intron lariat as well as to the release of a subset of snRNPs. In the second transesterification step, the 3' OH group of the upstream exon attacks the 3´ splice site, producing a fusion of the adjacent exons (spliced mRNA). The excised intron is subsequently degraded (Kornblihtt et al., 2013).

2.11 Alternative splicing and transcription

Similarly to other pre-mRNA processing events, splicing is coupled to transcription, in a way that both processes influence each other via co-ordinated mechanisms. One of those mechanisms is kinetic coupling, where the rate of Pol II-mediated elongation affects the pace at which splice sites and regulatory sequences appear in the nascent pre-mRNA during transcription. Sequences associated with Pol II pausing, promote retention of exonic cassettes in the mRNA (Nogue et al, 2002) whereas elongation-promoting factors increase Pol II speed and hence allow exon skipping (Roberts et al, 1998). Indeed, the idea of a slow Pol II facilitating inclusion of alternative exons is supported by global analyses where dozens of alternative splicing events are affected by the treatment of human cells with strong inhibitors of elongating Pol II (Ip, Schmidt, & Pan, 2011). Nevertheless, he effects of kinetic coupling on spicing are highly dependent on the microenvironment as well as the different combinations of splicing regulators.

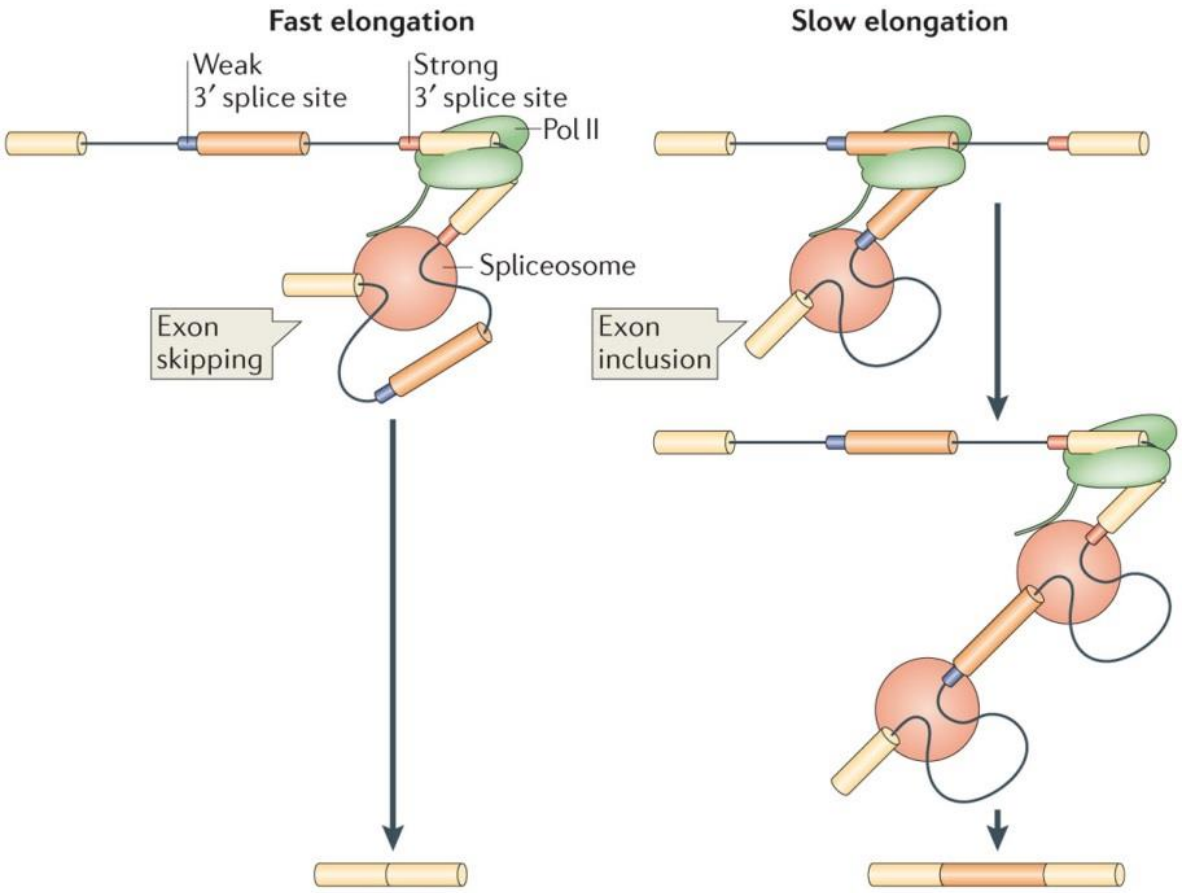


Figure 3: Effect of transcriptional elongation on alternative splicing. During RNA polymerase II (Pol II)-mediated transcription, exon skipping is achieved via fast elongation (left) that aids the recruitment of the spliceosome to the strong 3´ splice site of a downstream intron instead of the weak 3´ splice site of the upstream intron. On the other hand, slow elongation (right) favours the recruitment of spliceosome components to the upstream intron, which results in splicing commitment and exon inclusion (Kornblihtt et al., 2013).

2.12 Chromatin and alternative splicing

Over the years, increasing evidence pinpoints a key role of chromatin in alternative splicing of transcripts via the establishment of histone modifications and nucleosome positioning. The initial idea of chromatin structure affecting alternative splicing originates from studies in mammalian cells, where exonic inclusion was increased in response to DNA replication (Kadener et al., 2001). A more compact chromatin state would act as a barrier to Pol II elongation and promote high exonic inclusion as result of kinetic coupling.

Histone post-translational modifications, associating with either active (H3K36me3, H3K4me2/3, and H3K9Ac) or repressed (H3K9me2/3, and H3K27me3) transcription, are known to be indispensable regulators of alternative splicing. In neuronal cells, membrane depolarisation leads to accumulation of intragenic H3K9Ac which subsequently promotes an open chromatin state, increased Pol II elongation, and as a result exonic skipping in the *NCAM* (neural cell adhesion molecule) gene. On the contrary, during neuronal differentiation, the repressive H3K9me2 and H3K27me3 marks are highly enriched over the *NCAM* gene body, thus promoting exonic inclusion via compromised Pol II elongation (Schor et al, 2009) (Fig 4).

Taken together, external signals and differentiation pathways are capable of altering chromatin structure and ultimately influence the alternative splicing of mRNA transcripts.

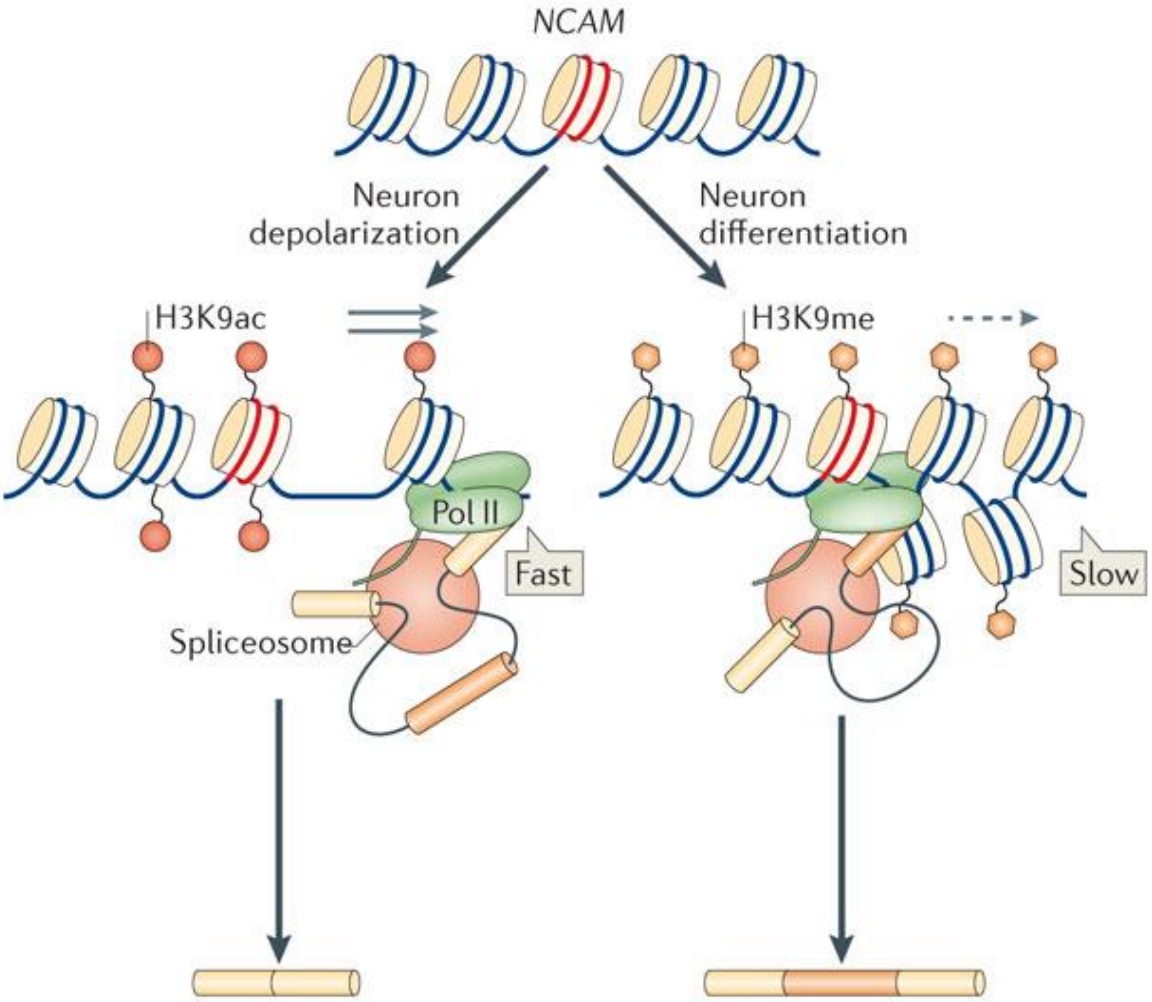


Figure 4: Effect of chromatin on alternative splicing. External stimuli as well as the differentiation state of the cell affect the chromatin state which in turn impacts alternative splicing decisions. Here, it is shown a collaborative mechanism between histone modifications and transcription with regards to kinetic coupling and splicing (Schor et al, 2009).

3 Aims

The function of FACT has been well characterised in both genetics and biochemical terms. However, its molecular functions in higher eukaryotes with regard to controlling transcriptional elongation and nucleosome occupancy are yet to be determined. Moreover, the means by which pluripotency is maintained in the presence of FACT are still unknown. In this project, we are trying to answer the following questions that will shed light to the molecular functions of FACT in mouse embryonic stem cells (mESCs);

1) What are the putative DNA binding sites for both FACT subunits (SSRP1, SUPT16)? Do they have autonomous and/or distinctive function?
2) How does FACT affect the transcriptional landscape?
3) Can the above changes in the transcriptome be reflected at a proteomic level? Does FACT serve as a docking site for additional proteins/TFs?
4) How is FACT involved in the integrity of the nucleosomal landscape?
5) Is FACT a critical component in the production of nascent RNA and the productive elongation of Pol II? To what extent do nucleosomes affect the progression of the FACT-Pol II complex?
6) Wrapping up all the above, can FACT maintain pluripotency?

4 Results & Discussion

4.1 FACT correlates with active gene expression

High expression of FACT is associated with stem or less-differentiated cells (Garcia et al., 2011). Indeed, we were able to confirm that diminishing levels of FACT correlate with the differentiation state of the cell. The highest FACT mRNA levels were observed for cancer and stem cell lines, whereas the lowest were observed for the potently differentiated ones (NIH3T3 & MEFs, **Fig. 5a**). We also employed RNA-seq data of ES to cardiomycte differentiation at different timepoints to assess FACT levels (Wamstad et al., 2012). Indeed, both levels of *Ssrp1* and *Supt16* are reduced with the priming of ES cells towards a differentiated state (**Fig. 5b**) Thus, we chose to explore the means by which FACT maintains an undifferentiated state.

Initially, we applied to mESCs a chromatin immunoprecipitation and sequencing (ChIP-seq) assay to identify potential DNA binding regions for both FACT subunits. Both SSRP1 and SUPT16 proteins displayed high similarity in binding capacity with the highest correlation observed over the gene body area of all genes (Pearson's correlation = 0.96, **Fig. 6a**), thus confirming their association to each other with regard to forming a protein complex.

Subsequently, we examined FACT co-enrichment with several other transcription factors, histone marks, and chromatin remodellers over the gene body area of all annotated genes (n = 13,348). High correlation scores were observed between SSRP1, SUPT16, H3K4me3, H3K27ac, and Pol II variants (Pol II S5ph, Pol II S2ph) confirming the role of FACT in active gene expression (**Fig. 6d & 7a**). A good correlation was also observed between both FACT subunits and Chd1, in

line with data demonstrating physical interaction and co-localization in mammalian cells (Kelley, Stokes, & Perry, 1999). However, a lower correlation was observed between FACT and H3K36me3 on a genome wide level. H3K36me3 has been shown to be able to directly recruit FACT to actively transcribed genes (Carvalho et al., 2013). We suspect that the enrichment of FACT subunits around the TSS might mask this potential correlation as FACT subunits also co-localise to the gene body of actively transcribed genes and enrich towards the TES, similarly to H3M36me3. Pearson's correlation remained elevated when we focused on active promoter and enhancer regions (*n* = 52,329) (**Fig. 7b**). Both subunits displayed very similar binding pattern to each other over the transcription start site (TSS), gene body and transcription end site (TES) of all the annotated genes (**Fig. 6b**) and were tightly linked to H3K4me3 (**Fig. 6c**). Importantly, FACT and Chd2 (functional equivalence to yeast Chd1 (de Dieuleveult et al., 2016)) exhibited opposite binding distribution patterns with the latter strictly encompassing the whole transcription unit (**Fig. 3c**). This is in agreement with genetic results obtained in yeast, demonstrating that Chd1 and yFACT have opposing functions during transcriptional initiation (Biswas et al., 2008). Our data suggest that FACT is enriched on active transcription sites and is mainly involved in transcriptional elongation and termination.

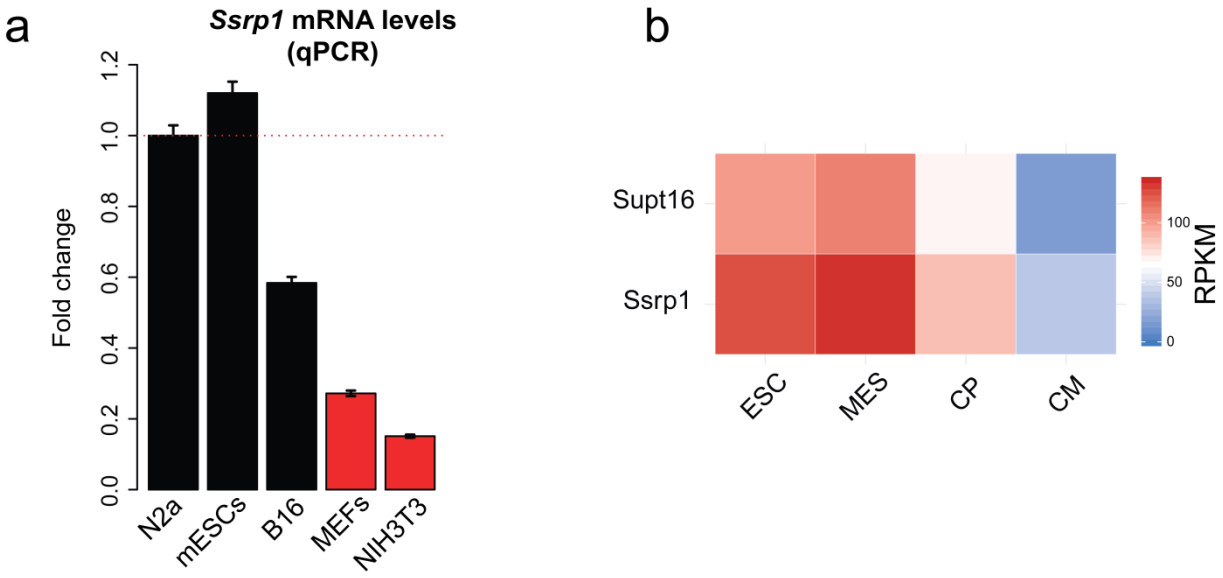


Figure 5: FACT levels are diminishing in a differentiated state. a, qPCR measuring *Ssrp1* mRNA levels among different cell lines. FACT levels are high in stem (mESCs) and cancer (N2a, B16) cells and are dramatically reduced in differentiated cells (MEFs, NIH3T3). **b**, Heatmap assessing the mRNA levels (RPKM) of FACT (*Ssrp1, Supt16*) at different timepoints of differentiation of ES cells to cardiomyocytes (Wamstad et al., 2012) (ESC = Embryonic Stem Cells, MES = Mesodermal cells, CP = Cardiac Precursors, CM = Cardiomyocytes).

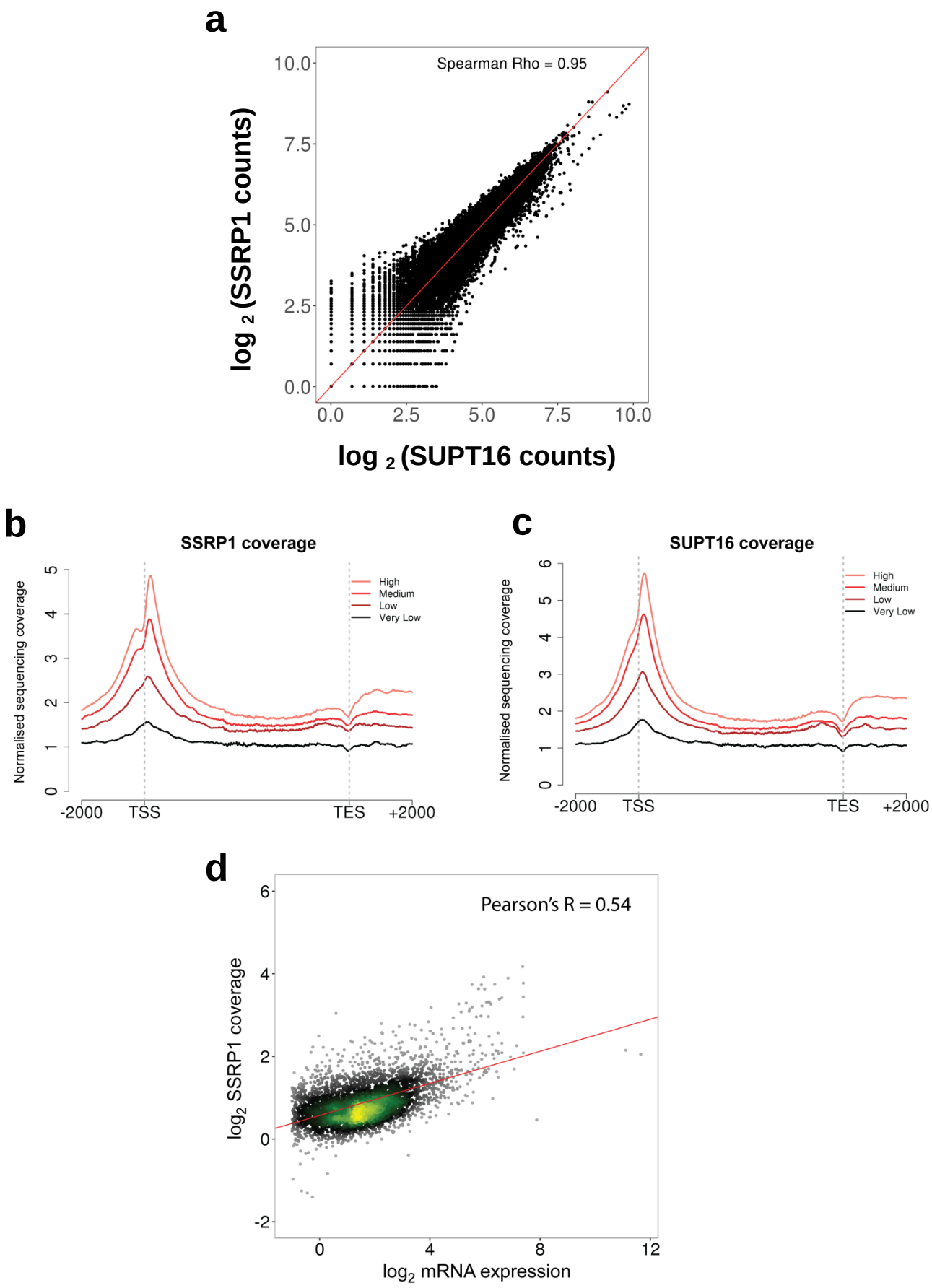

a
Spearman Rho = 0.95
log 2 (SSRP1 counts)
log 2 (SUPT16 counts)
b
SSRP1 coverage
Normalised sequencing coverage
High
Medium
Low
Very Low
-2000
TSS
TES
+2000
c
SUPT16 coverage
d
Pearson's R = 0.54
log2 SSRP1 coverage
log2 mRNA expression

Figure 6: **FACT correlates with active gene expression. a,** Scatterplot of SSRP1 and SUPT16 coverage over all annotated genes (n = 13,348). **b**, Distribution of SSRP1 relative to the TSS (± 2000 bp) and the TES (± 2000 bp) for four different gene classes ranked by level of RNA abundance (High, Medium, Low, Very Low). Gene clustering arises from division of mRNA expression into four different quantiles. **c**, Same as **b** but for SUPT16. **d**, Scatterplot of the $\log_2$SSRP1 coverage over $\log_2$ mRNA expression.

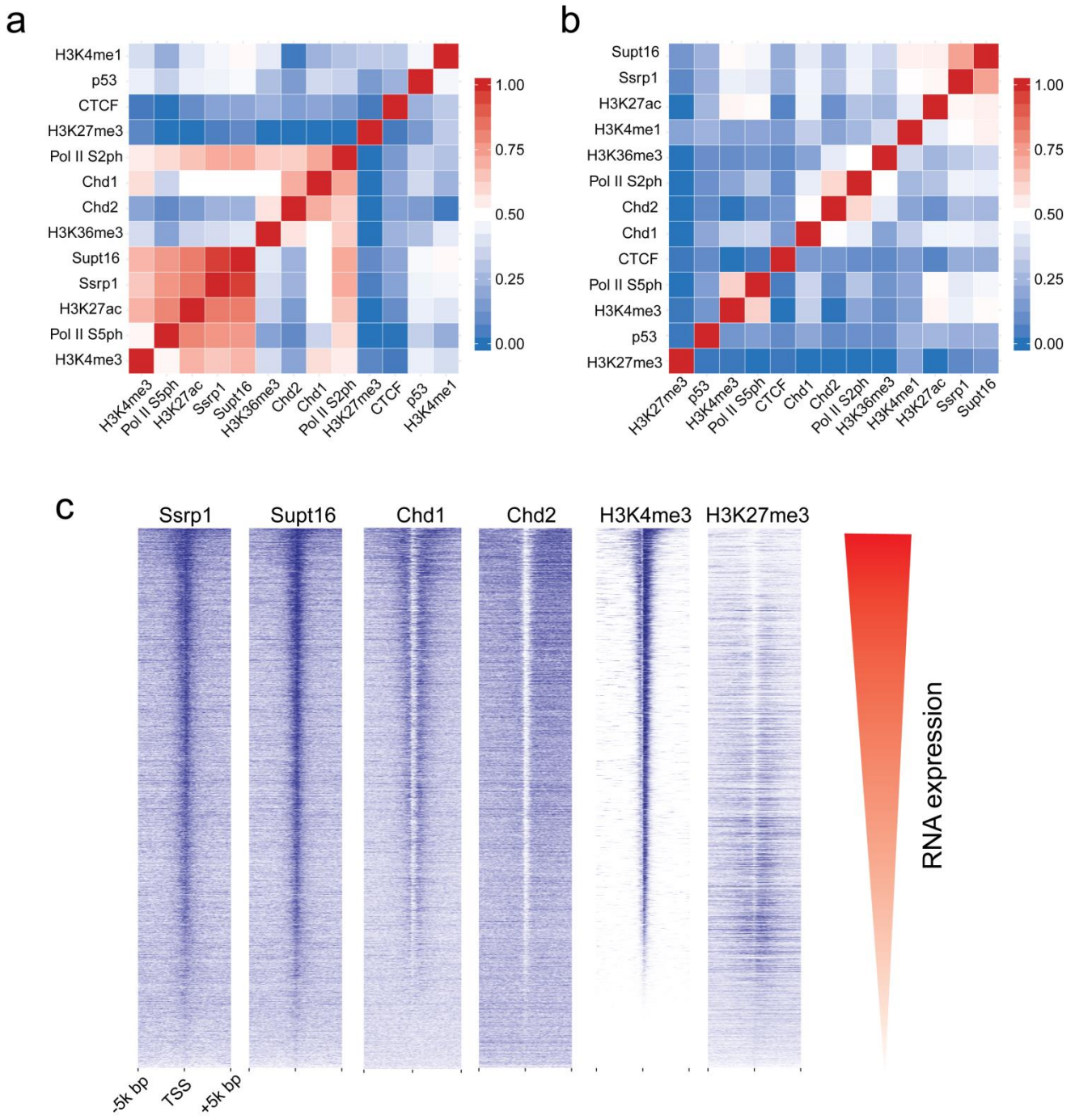


Figure 7: **Correlated occupancies across FACT-bound regions**. **a**, Heatmap representing Pearson's correlation between FACT subunits (SSRP1, SUPT16), and other factors over the gene body area of all unique annotated genes (*n* = 13,348). **b**, Same as **a** but for promoter/enhancer regions (*n* = 52,329) characterised by H3K27ac and/or H3K4me1 marks. **c,** Distribution of FACT and other factors (ChIP-seq tags indicated in blue) over the TSS of 13,348 unique RefSeq genes, sorted by H3K4me3 levels. Coinciding RNA expression levels are shown in red.

4.2 FACT modulates cell proliferation in ES cells

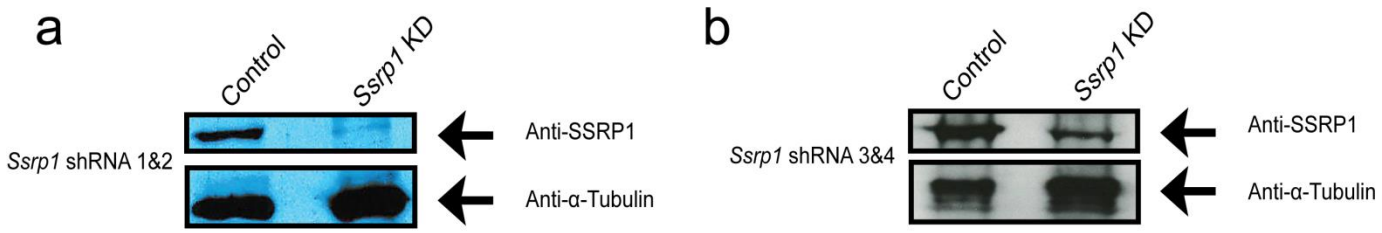
To investigate how FACT orchestrates transcriptional regulation in ESCs, we depleted SSRP1 levels using different combinations of short hairpin RNAs (shRNA – **Fig. 8a & 8b**). Surprisingly, we observed an increase in mESC proliferation following *Ssrp1* knock-down (KD) as measured by proliferation rate using MTT assays (**Fig. 8c & 8d**). However, ablation of *Ssrp1* levels in a neuroblastoma (N2a) and a skin cancer (B16) cell line did not cause a significant change in the proliferation rate of cells (**Fig. 9a**). In yeast, *Spt16* inactivation leads to lethality of the mutant strains (Biswas et al., 2005). Concomitant to higher eukaryotes, inhibition of FACT has been linked to tumour size reduction and diminished tumour progression (Gasparian et al., 2011; Koman et al., 2012). In addition, *Ssrp1 -/-* mouse embryos die soon after preimplantation (Cao et al., 2003). However, increased proliferation of mESCs following FACT depletion is in agreement with previously published data demonstrating that deletion of *Ssrp1* in Drosophila neuroblasts (embryonic cells giving rise to neuronal fibres) leads to increased proliferation of this cell type (Neumüller et al., 2011). Therefore, our data suggest a distinctive function of FACT in ES cells compared to cancer and yeast with regard to controlling cell proliferation.

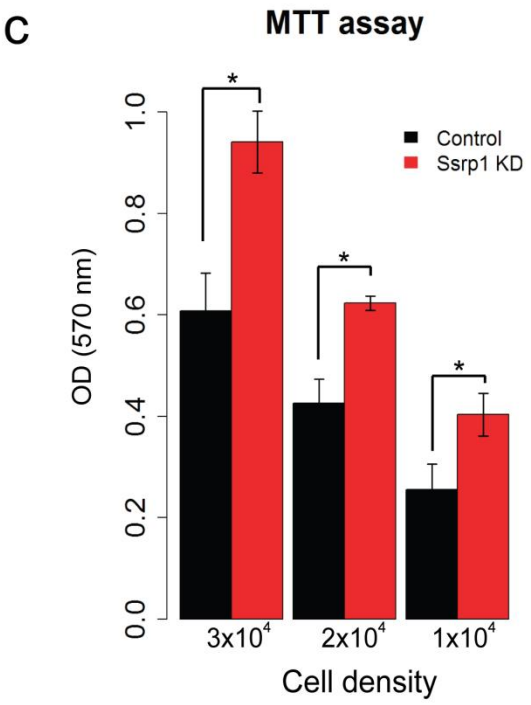
To understand the impact of FACT ablation on the transcriptional landscape, we sequenced the whole transcriptome (RNA-seq). In total, we characterised 3,003 differentially expressed genes; 1,655 down-regulated and 1,348 up-regulated (**Fig. 8e**). Down-regulated genes were over-represented for pathways involved in development, while up-regulated genes were involved in metabolic processes

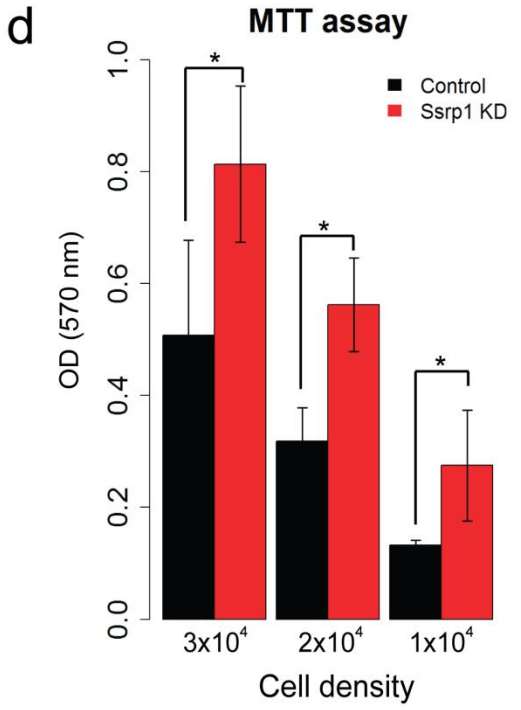
and positive regulation of cell growth (**Fig. 8f**), indicating that FACT negatively controls cell proliferation in ES cells.

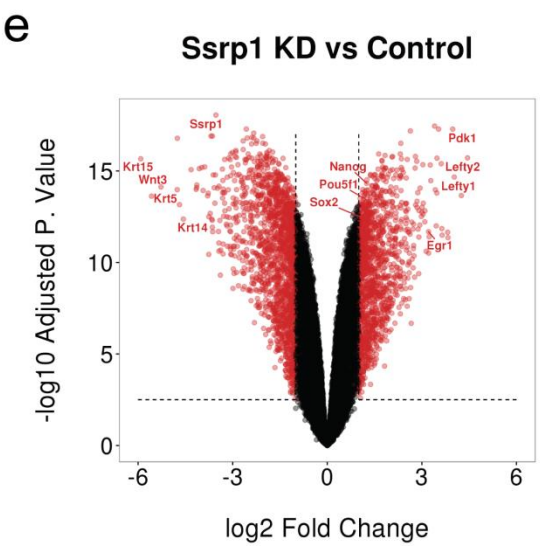
In addition, Mass spectrometry analysis of whole proteome between Control and *Ssrp1* depleted conditions, displayed a high correlation (Pearson R= 0.52) among mRNA and protein levels (**Fig. 9b**). Interestingly, the SUPT16 protein levels, but not the mRNA levels, decrease dramatically following depletion of *Ssrp1* levels, suggesting that the FACT complex is functional and able to exert its function only when both subunits are present.

This interdependence has also been previously reported in human cells where *in vitro* differentiation (Garcia et al., 2011) or depletion (Safina et al., 2013) of either subunit leads to deterioration of both subunits. In addition, they similarly report no change in gene expression levels of either subunit upon FACT depletion. These data support a model where a very stable complex is promptly disrupted in specific conditions (e.g. cell differentiation) when it is no longer required. Moreover, this inter-regulation of SSRP1 and SUPT16 levels provides a FACT reformation barrier that is disrupted in abnormal conditions such as cancer.


a
Control
Ssrp1 KD
Ssrp1 shRNA 1&2
Anti-SSRP1
Anti-α-Tubulin
b
Control
Ssrp1 KD
Ssrp1 shRNA 3&4
Anti-SSRP1
Anti-α-Tubulin


c
MTT assay
Control
Ssrp1 KD
OD (570 nm)
0.0
0.2
0.4
0.6
0.8
1.0
*
*
*
3x10^4
2x10^4
1x10^4
Cell density


d
MTT assay
Control
Ssrp1 KD
OD (570 nm)
0.0
0.2
0.4
0.6
0.8
1.0
*
*
*
3x10^4
2x10^4
1x10^4
Cell density


e
Ssrp1 KD vs Control
-log10 Adjusted P. Value
0
5
10
15
Ssrp1
Krt15
Wnt3
Krt5
Krt14
Nanog
Pou5f1
Sox2
Pdk1
Lefty2
Lefty1
Egr1
-6
-3
0
3
6
log2 Fold Change

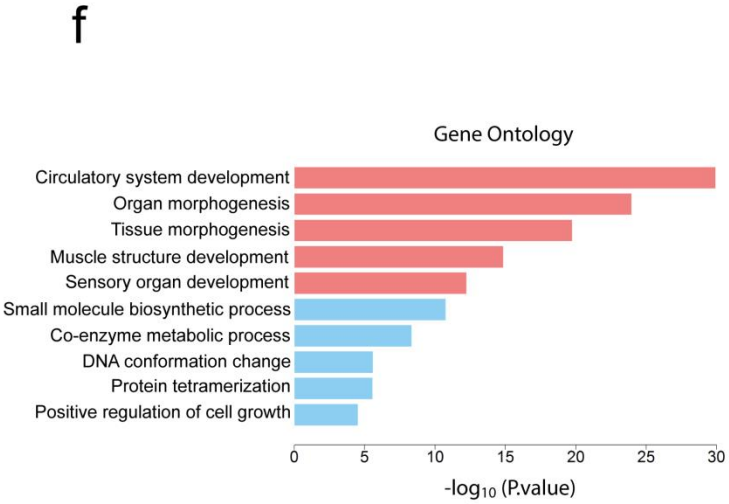

f
Gene Ontology
Circulatory system development
Organ morphogenesis
Tissue morphogenesis
Muscle structure development
Sensory organ development
Small molecule biosynthetic process
Co-enzyme metabolic process
DNA conformation change
Protein tetramerization
Positive regulation of cell growth
0
5
10
15
20
25
30
-log10 (P.value)

Figure 8: **FACT controls cell proliferation and developmental pathways**. **a**, Western blots after transfection of mESCs with different combinations of *Ssrp1* shRNA vectors (shRNA 1&2). Anti-α-Tubulin was used as a reference. **b**, Same as **a** but for a different combination of *Ssrp1* shRNA vectors (shRNA 3&4). **c**, MTT assay following transfection with *Ssrp1* shRNA 1&2 vectors. Values are mean and SE of three independent transfection experiments are displayed. Significance was calculated via a two-tailed *t-test* (*$P < 0.05$). **d**, Same as **c** but for a different combination of *Ssrp1* shRNA vectors (shRNA 3&4). **e**, Volcano plot of differentially expressed genes between the Control and KD group. Values with logFC > 1 or logFC <-1 and Adjusted P.value < 0.01 are highlighted in red. **f**, Gene ontology analysis of all differentially expressed genes (Red: pathways for down-regulated genes, Blue: pathways for up-regulated genes).

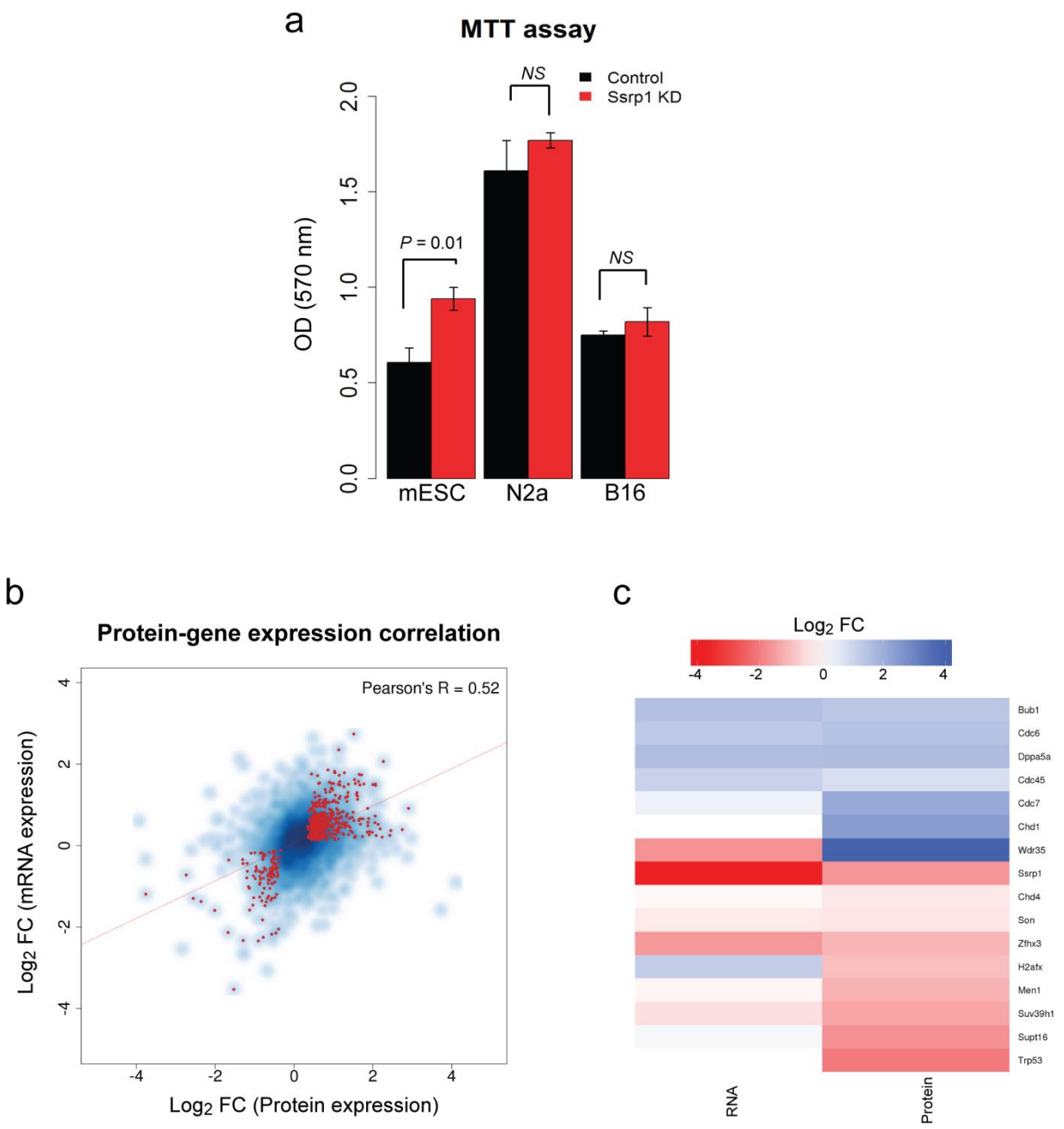


Figure 9: **Selective function of FACT in controlling mRNA and protein levels in ES cells**. **a**, MTT assay following transfection with *Ssrp1* shRNA 1&2 vectors in ES, N2a, and B16 cells. Values are mean and SE of three independent transfection experiments are displayed. Significance was calculated via a two-tailed *t-test*. **b,** Scatteplot of logFC of mRNA and protein levels following depletion of *Ssrp1.* **c**, Heatmap displaying the logFC (mRNA and protein) of several genes after depletion of *Ssrp1* levels.

4.3 FACT controls expression of active genes via nucleosome deposition

Despite FACT occupancy exhibiting high correlation with active gene expression, depletion of FACT levels in mESCs causes an abrupt de-regulation of genes that are bound by it. A low correlation (Pearson's R = 0.11) was observed between the coverage of SSRP1 (ChIP-seq) and the gene fold change (RNA-seq) of those genes in the *Ssrp1* KD (**Fig. 10a**). Apart from down-regulated genes, we discovered a class of genes whose expression was not influenced by FACT depletion despite being a FACT-target. Interestingly, a third class of FACT-bound genes emerged, which displayed high levels of RNA, suggesting that FACT negatively regulates the expression of these genes in mESCs by preventing over-expression.

Since FACT is responsible for the remodelling of nucleosomes in the front of RNA polymerase and the re-establishment of nucleosome integrity in its wake (Formosa, 2012), we next sought to investigate the changes in the nucleosomal landscape triggered by the absence of FACT. MNase-resistant mononucleosome-sized DNA fragments (135-170 bp) were purified from control and *Ssrp1*-depleted conditions, and sequenced. Nucleosome occupancy was plotted for six different gene classes according to the presence of SSRP1 in the control group (ChIP-seq) and their relative Fold Change in the *Ssrp1* KD state (RNA-seq) (**Fig. 10b**). Genes whose expression was not impacted in the KD ("No change" class) displayed similar mononucleosome patterns among SSRP1 and Non-SSRP1 targets, implying that FACT affects neither nucleosome occupancy nor gene expression in this gene class. On the other hand, genes that are down-regulated in the *Ssrp1* KD ("Down-regulated" class) and bound by FACT exhibit a global mononucleosomal shift by a few nucleotides right after the +1 nucleosome. In contrast, genes which are bound by FACT and up-regulated in the *Ssrp1* KD

("Up-regulated" class) display two mononucleosomes in the Nucleosome Free Region (NFR), which are lost in the absence of FACT. In addition, the +2 and +3 nucleosome occupancy drops in the absence of FACT. This increased nucleosome occupancy at the promoter region is highly reproducible among the different replicates and seems to arise from three distinctive gene clusters of diverse expression levels (**Fig. 11**). In concordance to the RNA-seq data, FACT dampens the expression of pro-proliferative genes by regulating the positioning of nucleosomes in the vicinity of the TSS.

Corroboratively, this is in agreement with data generated in *S. cerevisiae* where FACT regulates local nucleosomal stability to maintain the repression of non-coding transcripts (Feng et al., 2016). Moreover, at the long non-coding *SER3* locus in *S. cerevisiae*, absence of FACT leads to loss of nucleosomes at the promoter, which subsequently triggers expression of the gene (Hainer, Pruneski, Mitchell, Monteverde, & Martens, 2011), indicating that FACT might regulate expression of genes indirectly through the positioning of nucleosomes.

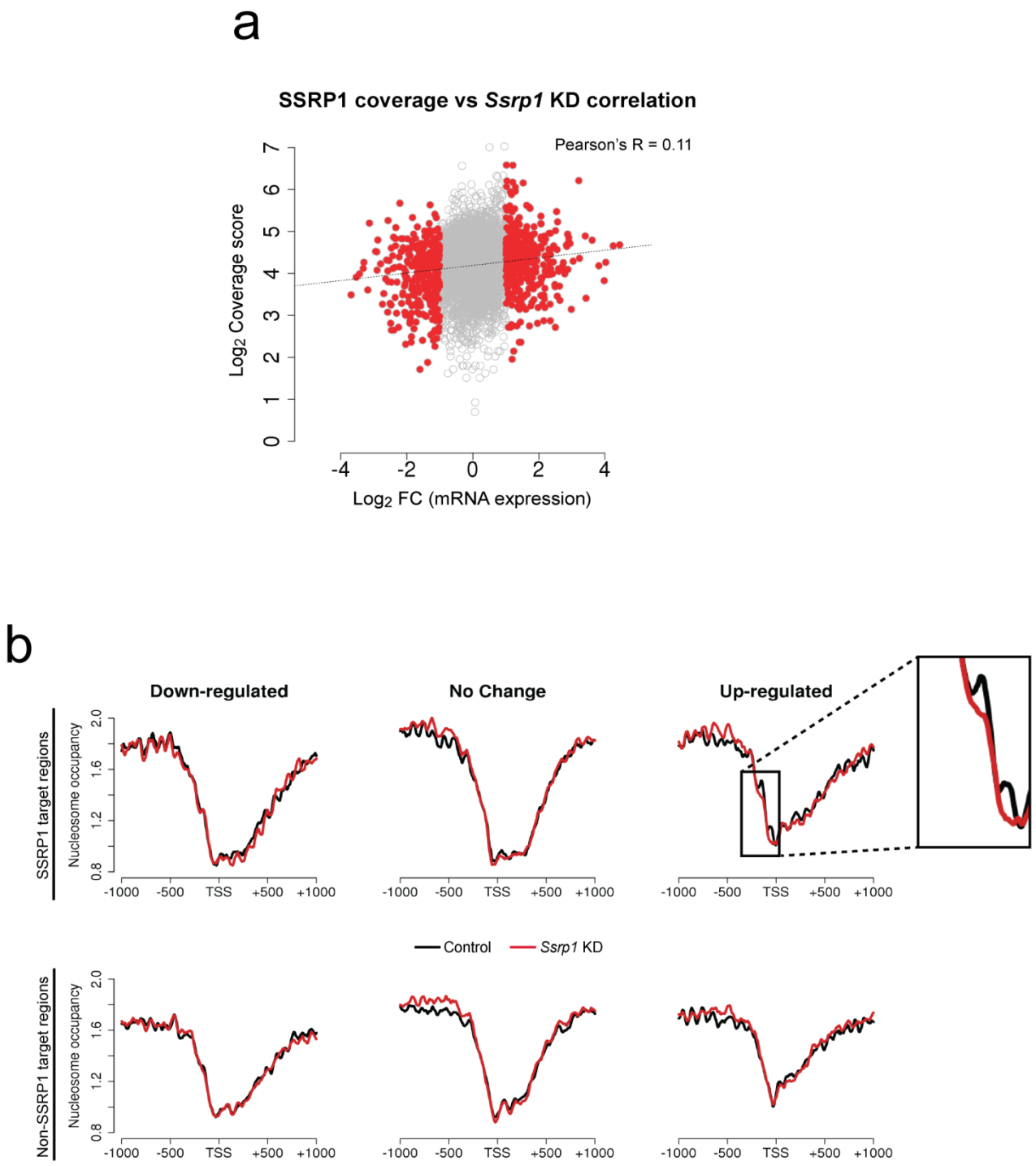


Figure 10: **Nucleosome deposition by FACT dampens gene expression.** **a**, Scatterplot of log (SSRP1 coverage) (ChIP-seq) over logFC (RNA-seq). **b**, Nucleosome occupancy of all genes indicated in **a**. Datasets are split by their FACT occupancy status (SSRP1 and Non-SSRP1 targets) and their relative transcriptional direction ("Down-regulated", "No Change", "Up-regulated").

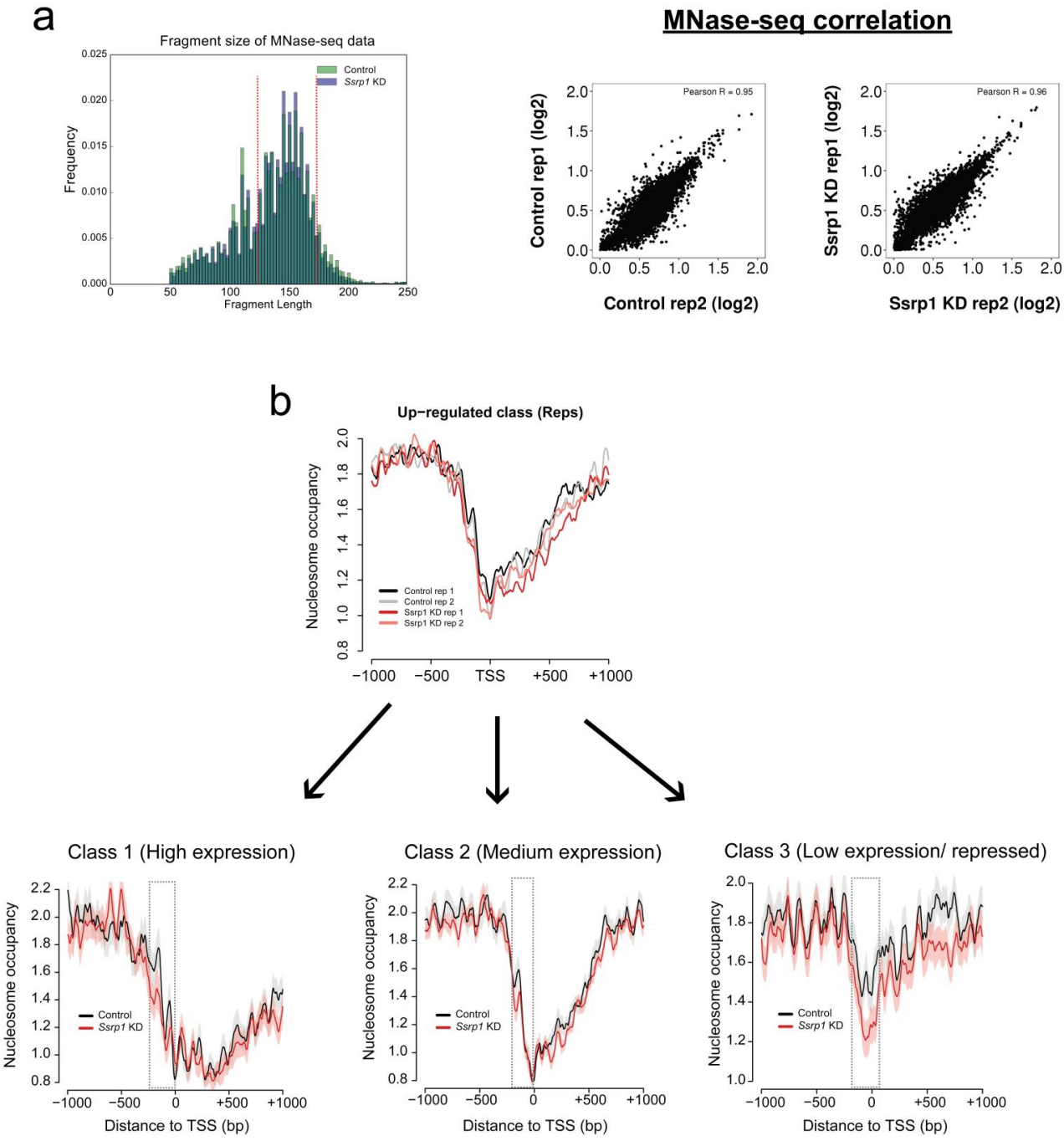


Figure 11: **Reproducibility assessment of MNase-seq dataset a,** Histogram showing the sequenced paired-end fragments lengths extracted from Control and *Ssrp1* depleted conditions. Fragments between 135-170 bp (indicated in red) have been computationally selected and used to plot mono-nucleosomal occupancy over promoter regions. Correlation scatterplots (MNase-seq) accessing replicate reproducibility in each condition. Pearson's correlation is indicated at the top of each plot. **b,** Nucleosome occupancy metaplots for each replicate for the composite metaplot in **Figure 10b**. This nucleosome occupancy at the promoter region derives from three distinctive gene clusters of diverse gene expression.

4.4 FACT and chromatin remodellers control gene expression

Next, we wanted to determine whether there are factors that are decisive in determining if a gene changes its expression level upon depletion of FACT. We turned to nucleosome remodellers, and examined their distribution around the TSS of all three gene classes (SSRP1 targets). No change was observed in occupancy for Chd1, Chd2, or Chd4, but we detected a striking difference in Ep400, Smarcad1, and Chd6/8/9 (**Fig. 12**). This implicates that changes in the presence and occupancy degree of the latter ones are decisive for transcriptional activity upon depletion of FACT levels.

The presence of FACT elicits a tight co-ordination of transcription over H3K4me3 and bivalent (H3K4me3 and H3K27me3) promoters. In a highly bivalent state, Ep400, Chd6, and Chd8 bind exclusively to the -1 nucleosome upstream of the TSS. We hypothesize that in the absence of FACT, recruitment of the SWI/SNF complex occurs over certain promoter regions and further represses expression of those lineage-specific genes (Wilson & Roberts, 2011). This speculation is reinforced by the fact that we observe a nucleosomal shift in the “Down-regulated” class (*Ssrp1* KD); a mechanistic feature that has been previously reported to be driven by SWI/SNF (Nocetti & Whitehouse, 2016). In highly active genes (H3K4me3) both the +1 and -1 nucleosomes are engaged by Ep400, Chd6, and Chd8. In addition, our data suggest that promoters that are occupied by FACT, enable Set1 binding further downstream of the TSS which subsequently allows broad H3K4me3 and Pol II distribution. Other remodellers might bind as well but their function and contribution to transcriptional regulation remains unclear.

Nevertheless, the presence of Chd9 on the +1 nucleosome seems to play an essential part in the transcriptional decision between “Up-regulation” and “No-change” when FACT is absent. Loss of FACT probably allows Chd9 to re-order chromatin and facilitate binding of TFs that will subsequently allow increased downstream mRNA production. Expression of Chd9 has been previously shown to display a critical role in cellular differentiation and bone growth (Marom, Shur, Hager, & Benayahu, 2006), whereas more recently a link between Chd9 and ribosomal transcription has been reported (Salomon-Kent et al., 2015). Considering that the high H3K4me3 “Up-regulated” class exhibits elevated occupancy of Chd9 on the +1 nucleosome, we hypothesise that Chd9 reorganises chromatin and allows binding of GTFs in the absence of FACT.

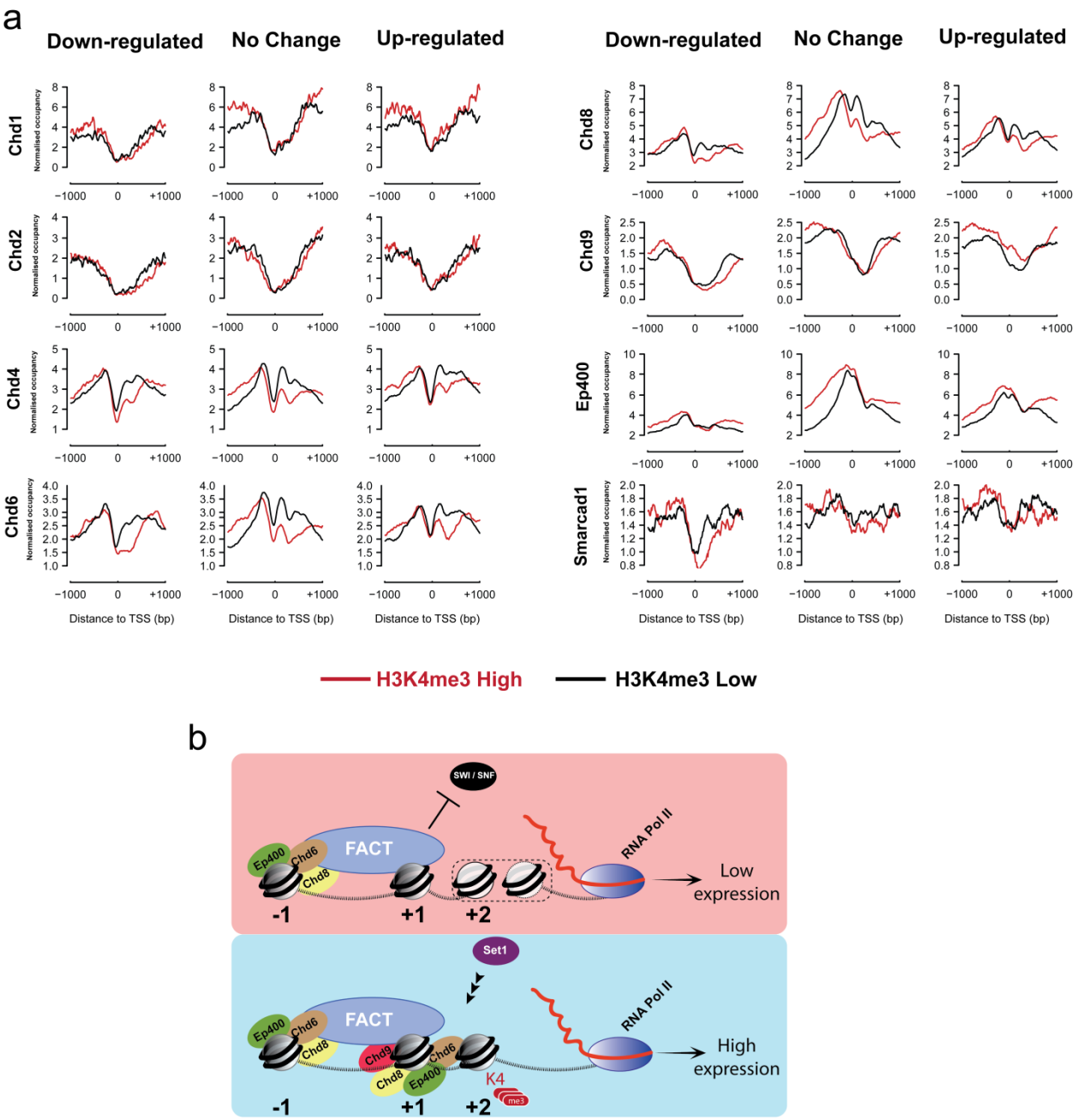


Figure 12**: Distribution of different nucleosome remodellers over the TSS of FACT-bound promoters**. Datasets are split by their relative chromatin (H3K4me3 High, H3K4me3 Low) and gene expression status ("Down-regulated", "No-Change", "Up-regulated"). Model of how FACT and chromatin remodellers might regulate transcriptional directionality over lowly (1st model) and highly (2nd model) expressed promoters. The whole model was derived by comparing the H3K4me3 High class

between the “Up-regulated” (High expression) and “Down-regulated” (Low expression) gene clusters. In the 1st model, the presence of FACT will obstruct binding of the SWI/SNF complex (**Fig. 19c**-Smarcad1) to prevent down-regulation of lineage-specific genes. Moreover, global nucleosomal shifting after the +2 nucleosome will be avoided. In the 2nd model, the presence of FACT enables binding of Set1 and ultimately establishment of a broad H3K4me3 mark. In addition, the Ep400-Chd6-Chd8-Chd9 complex in the +1 nucleosome probably marks those promoters for up-regulation in the absence of FACT.

4.5 Nucleosome deposition affects RNA pol II pausing index

Over the last decade it has become apparent that promoter proximal pausing of RNA Pol II plays an important role in regulating gene expression (I Jonkers & Lis, 2015). To explicate the interplay between the Pol II – FACT complex and RNA Pol II transcription, we performed NET-seq (Mayer et al., 2015) to identify transcribing RNA Pol II over SSRP1- and Non-SSRP1-bound regions. A higher correlation and a significantly higher slope of nascent RNA – mRNA expression was observed over the SSRP1-bound regions in both the Control and the *Ssrp1* KD state (**Fig. 13a,b**) suggesting diverse elongation rates of Pol II in the absence of FACT. To confirm this, we measured the travelling ratio of Pol II over the three different gene classes ("Down-regulated", "No Change", and "Up-regulated"). Compared to the Non-SSRP1 target regions, SSRP1-bound regions whose expression is either down-regulated or not changed, did not display a significant difference in the Pol II pausing index (**Fig. 13c**). Nevertheless, genes that were upregulated in the absence of FACT exhibited, in the Control state, a lower pausing index compared to their Non-SSRP1 target counterparts. However, this difference was preserved in the absence of FACT, indicating that FACT has no effect on the release of paused Pol II and its successful progression towards transcriptional elongation. Given the differences in Pol II S2 density profiles (ChIP-seq- **Fig. 13c**) and promoter architecture as defined by MNase-seq (**Fig. 10b**), it is tempting to speculate that genes upregulated by knock-down of FACT undergo a different pausing mechanism.

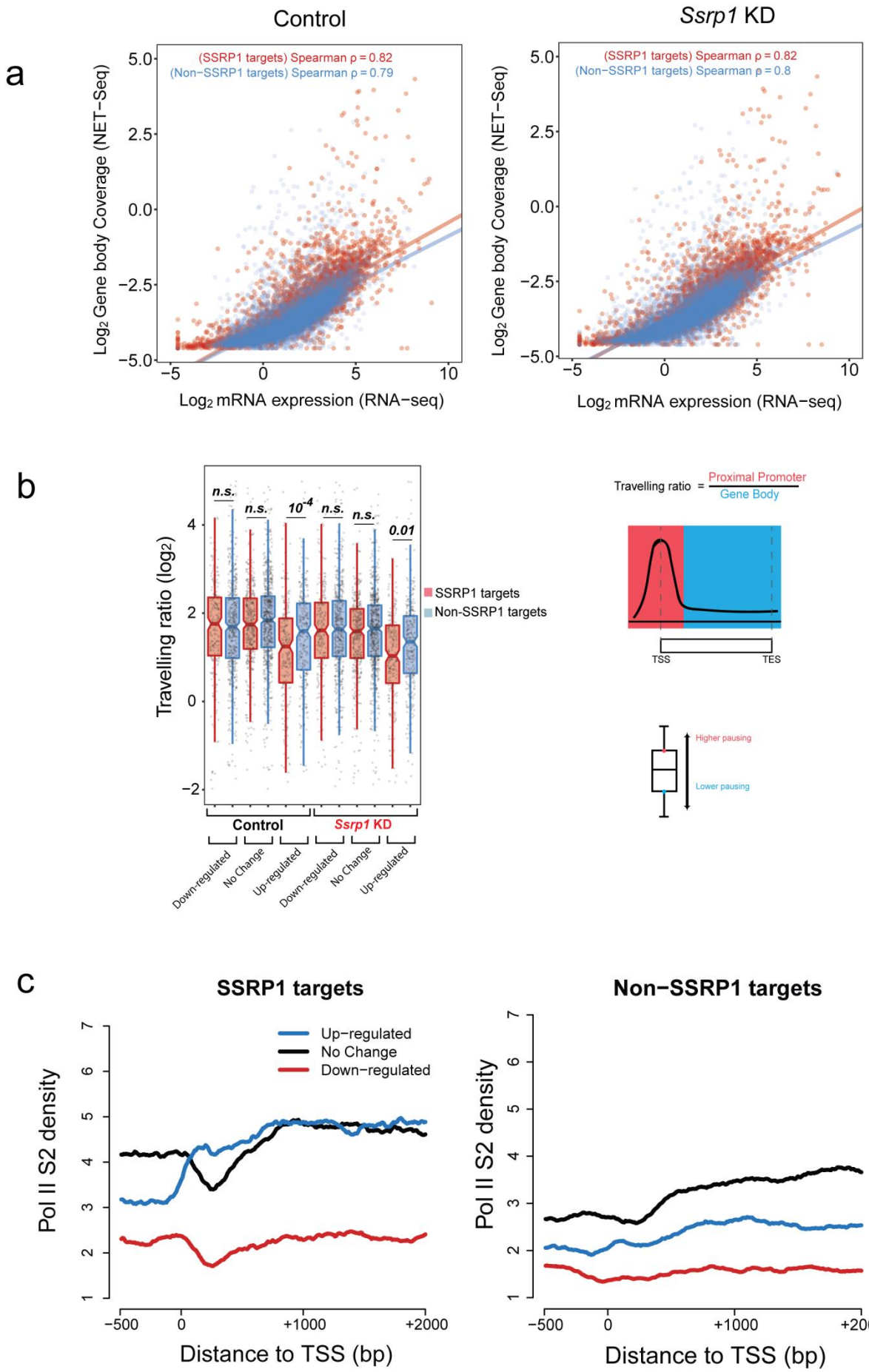

a
Control
Ssrp1 KD
(SSRP1 targets) Spearman ρ = 0.82
(Non−SSRP1 targets) Spearman ρ = 0.79
(SSRP1 targets) Spearman ρ = 0.82
(Non−SSRP1 targets) Spearman ρ = 0.8
Log2 Gene body Coverage (NET−Seq)
Log2 mRNA expression (RNA−seq)
b
Travelling ratio (log2)
n.s.
n.s.
10−4
n.s.
n.s.
0.01
SSRP1 targets
Non-SSRP1 targets
Control
Ssrp1 KD
Down-regulated
No Change
Up-regulated
Down-regulated
No Change
Up-regulated
Travelling ratio = Proximal Promoter / Gene Body
TSS
TES
Higher pausing
Lower pausing
c
SSRP1 targets
Non−SSRP1 targets
Up−regulated
No Change
Down−regulated
Pol II S2 density
Distance to TSS (bp)
−500
0
+1000
+2000

Figure 13: **Pol II elongation rate is altered in the absence of FACT. a**, Scatterplots of log gene body coverage (NET-seq) versus log mRNA expression (RNA-seq) for SSRP1 (n=4,576) and Non-SSRP1 (n=8,844) target regions in the Control (Z-score = 5.3, $P < 10^{-5}$) and *Ssrp1* KD (Z-score = 7.2, $P < 10^{-5}$) state. **b**, Measure of Pol II pausing. Travelling ratio is defined as NET-seq density of proximal promoter versus gene body area. Normalised travelling ratios for each gene class are displayed as boxplots. Densities in each condition have been normalised to "Up-regulated / Non-SSRP1 targets". The Wilcoxon rank test was used to calculate significance between SSRP1 and Non-SSRP1 targets. **c,** Distribution of elongating Pol II variant (S2 ph – WT ES cells) over the TSS (± 2000 bp) of SSRP1- and Non-SSRP1-bound regions split by transcriptional regulation ("Up-regulated", "No Change", "Down-regulated"). The "Up-regulated" SSRP1-bound gene class exhibits a different occupancy pattern.

4.6 Pol II density anti-correlates with nucleosome density

To further explicate the impact of nucleosome occupancy on Pol II density, we initially interrogated the nuclesomal landscape over CTCF and H3K27Ac regions. The reason behind this is the fact that we employed a different MNase-seq protocol that only retains soluble chromatin and thus avoids calculation of biases arising from formaldehyde cross-linking. We split the data to either small (< 80 bp) or larger (135-170 bp) fragments corresponding to transcription factor footprints and mono-nucleosomes, respectively. Nucleosomal occupancy over CTCF and H3K27Ac regions seems to be identical to previous reports (Carone et al., 2014; Teif et al., 2012), further reinforcing the legitimacy of our data (**Fig 14 a,b**). We next sought to determine nucleosomal distribution over all exonic regions. Data obtained from insoluble chromatin suggest that nucleosome occupancy in the intronic regions is higher compared to the exons, as opposed to the soluble chromatin where the opposite effect is observed (**Fig 14 c**). Our MNase-seq findings are corroborative to the soluble chromatin pattern (**Fig. 14 d**) and anti-correlate with Pol II pausing (**Fig. 14 e**). Increased mono-nucleosome occupancy of exons in the *Ssrp1* KD state negatively affects Pol II pausing over those regions. Therefore, our data suggest that the presence of nucleosomes influences to some extent pausing of Pol II over splice sites as well as over the whole exonic region.

In addition, our MNase-seq data are not in agreement with previous reports stating that nucleosomal occupancy of exons is higher compared to introns (Tilgner et al., 2009). We followed an MNase treatment protocol that gets rid of insoluble chromatin and retains soluble nucleosomes that are not bound by proteins/ transcription factors. This allows a direct interrogation of the soluble

nucleosomal landscape and makes it ideal for studying proteins involved in chromatin/ nucleosome remodelling. Higher intronic nucleosome occupancy can also be observed in other published data where soluble chromatin has been extracted (Carone et al., 2014). In addition, chemical mapping of nucleosomes also supports a lower occupancy of nucleosomes over exonic regions (Voong et al., 2016). Nevertheless, the ingenuity of our results can be further supported by the nucleosomal distribution adjacent to CTCF and H3K27Ac binding sites which is identical to other MNase-seq profiles obtained both from soluble and insoluble chromatin (Carone et al., 2014; Teif et al., 2012).

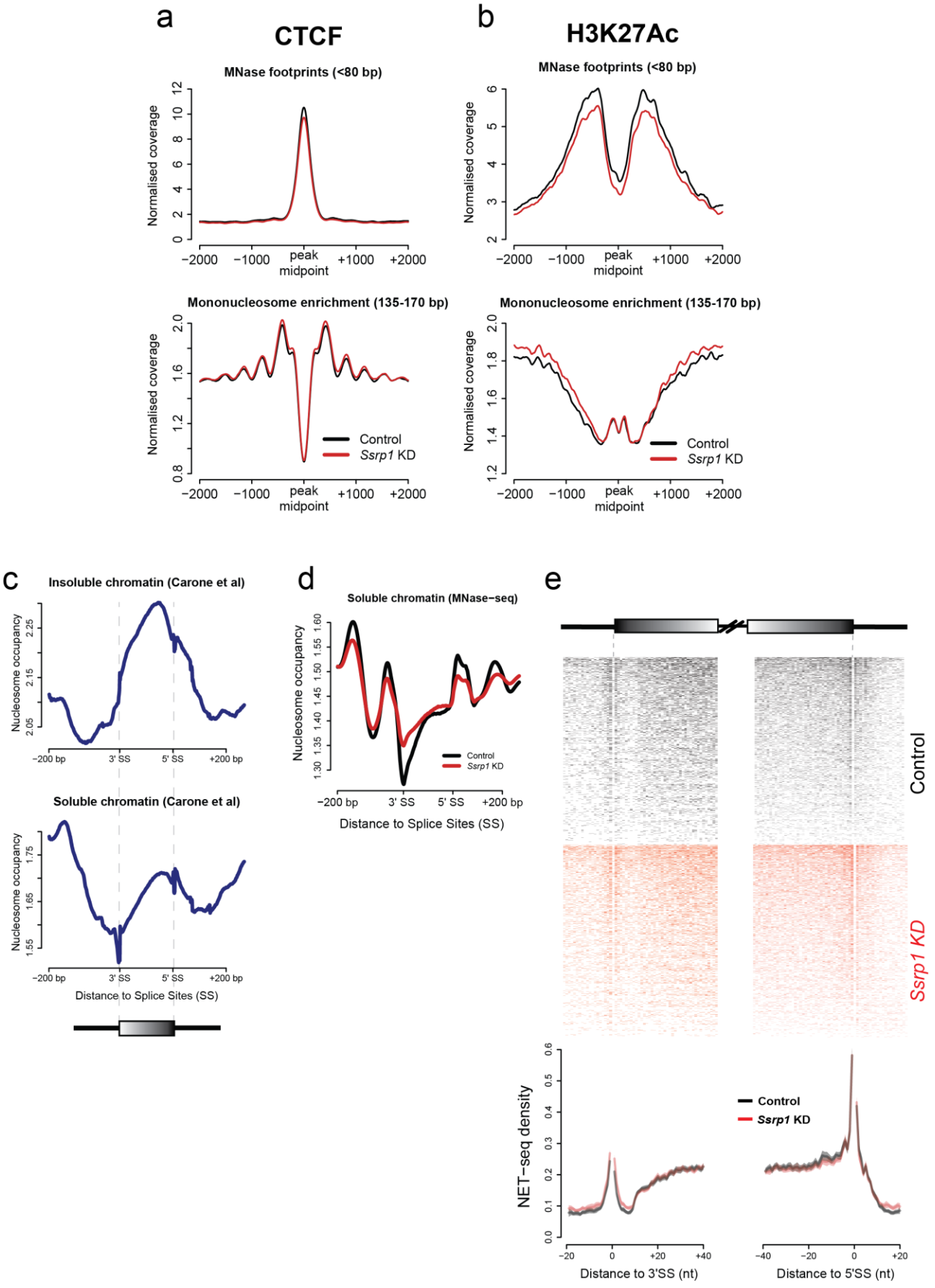

a
CTCF
MNase footprints (<80 bp)
Normalised coverage
−2000 −1000 peak midpoint +1000 +2000
Mononucleosome enrichment (135-170 bp)
Control
Ssrp1 KD
b
H3K27Ac
MNase footprints (<80 bp)
Mononucleosome enrichment (135-170 bp)
c
Insoluble chromatin (Carone et al)
Nucleosome occupancy
−200 bp 3' SS 5' SS +200 bp
Soluble chromatin (Carone et al)
Distance to Splice Sites (SS)
d
Soluble chromatin (MNase−seq)
Distance to Splice Sites (SS)
e
Control
Ssrp1 KD
NET−seq density
Distance to 3'SS (nt)
Distance to 5'SS (nt)

Figure 14: **Quality control assessing MNase-seq/ NET-seq integrity. a**, CTCF peak midpoints were used as a reference and nucleosome occupancy for short (<80 bp) and long (135-170 bp) MNase footprints was plotted over the TSS. **b**, Same as **a** but for H3K27Ac. Both CTCF and H3K27Ac profiles are consistent with previous studies(Carone et al., 2014; Teif et al., 2012). **c**, MNase-seq datasets of soluble (higher intron occupancy) and insoluble (higher exon occupancy) chromatin retrieved from Carone *et al.* **d**, Mean nucleosomal density (207,232 exons) of our insoluble MNase-seq dataset. The soluble nucleosome profile in **"c"** is highly consistent to our MNase treated samples for both conditions where nucleosomal occupancy on introns is similar or higher compared to the exons. Identical occupancy is also observed by chemical mapping of nucleosomes (Voong et al., 2016). **e**. NET-seq heatmaps and density plots over 41,356 exons with the highest Pol II coverage. Solid lines on the NET-seq meta-exon plots indicate the mean values, whereas the shading represents the 95% confidence interval. Density around the 3'SS and 5'SS was scaled independently. Pol II occupancy over exonic regions is higher compared to introns as opposed to nucleosome occupancy.

4.7 FACT enables broad distribution of H3K4me3

Taken into consideration the unanticipated MNase-resistance over the NFR region of the “Up-regulated” class in the control group, we sought to understand how chromatin modifications localise across the promoter regions of this class of genes. We compared regions that were characterised either by high or low levels of H3K4me3 (**Fig. 15a**). Intriguingly, we observed that the SSRP1-bound, high H3K4me3 regions exhibit a broad distribution of the chromatin mark; a result consistent with Set1 (responsible for H3K4me3 establishment) distribution over the TSS of the above genes. Set1 is recruited to the 5’ ends of genes by the Pol II S5ph variant while establishment of H3K4me3 enables Pol II elongation (Hsin & Manley, 2012). No broad H3K4me3 distributions were observed in the “No Change” or the “Down-regulated” gene classes (**Fig. 15 b,c**), with the latter being characterised by low gene expression and high enrichment of H3K27me3. Broad H3K4me3 domains have been previously linked before to cell identity and increased transcriptional elongation (Benayoun et al., 2014; Chen et al., 2015; Liu et al., 2016). Therefore, we focused more on the high H3K4me3 class of genes in the “Up-regulated” class in order to understand the consequences of directionality. Intriguingly, key pluripotency factors (e.g. Oct4,Sox2, Nanog, and Klf4) belong to this gene cluster.

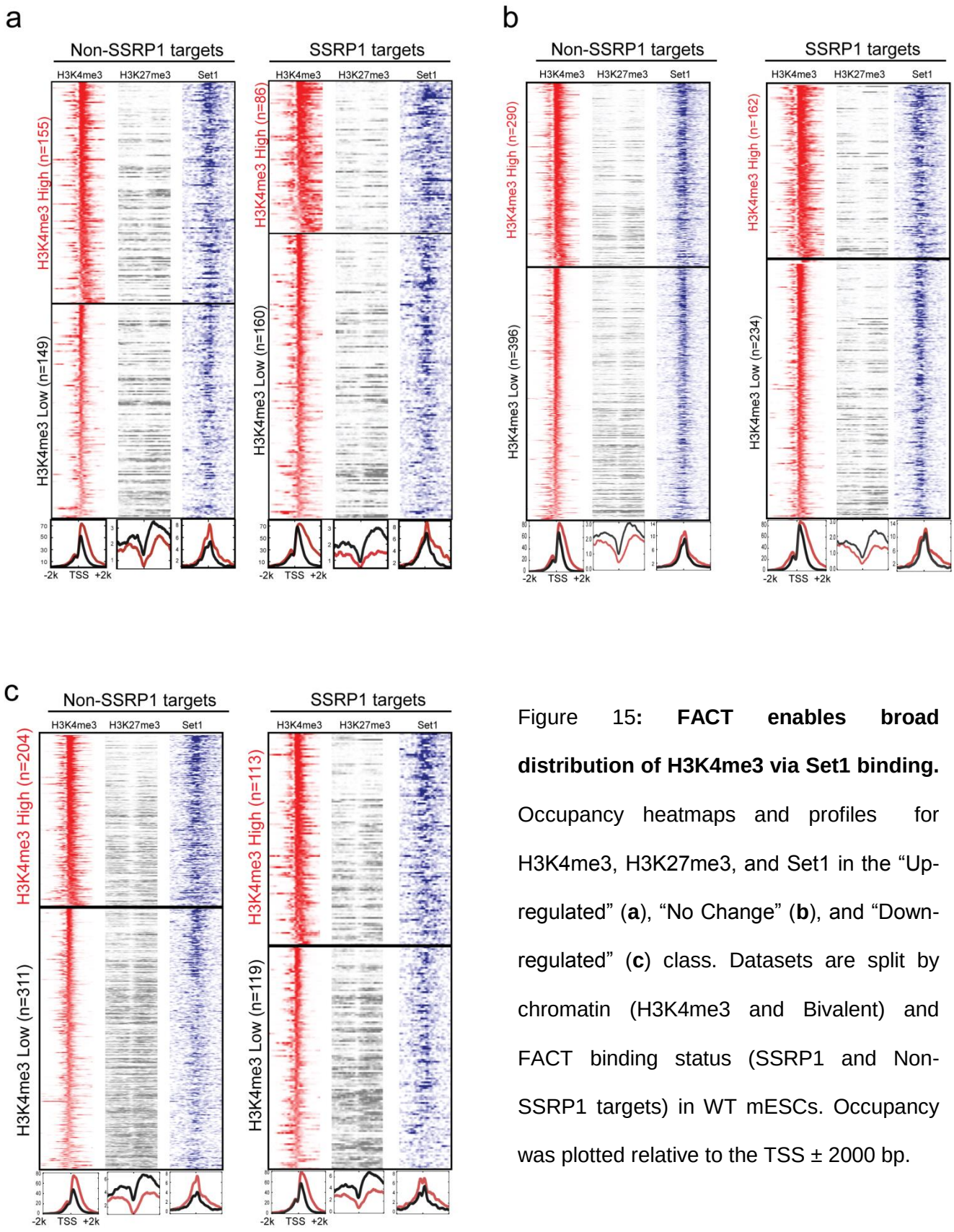


Figure 15**: FACT enables broad distribution of H3K4me3 via Set1 binding.** Occupancy heatmaps and profiles for H3K4me3, H3K27me3, and Set1 in the "Up-regulated" (**a**), "No Change" (**b**), and "Down-regulated" (**c**) class. Datasets are split by chromatin (H3K4me3 and Bivalent) and FACT binding status (SSRP1 and Non-SSRP1 targets) in WT mESCs. Occupancy was plotted relative to the TSS ± 2000 bp.

4.8 FACT restricts divergent Pol II via nucleosome deposition

H3K4me3 serves as a recruitment platform for TFs and RNA Pol II (Sims et al., 2007). Strikingly, NET-seq density plots identified that SSRP1 targets displayed high levels of promoter proximal pausing of RNA Pol II at the TSS, but also a strong and wide Pol II distribution similar to the H3K4me3 mark with little signal of divergent Pol II in the anti-sense strand (**Fig. 16**- H3K4me3 High- SSRP1 targets). The broad distribution in the sense strand was also confirmed by employing published GRO-seq data (**Fig. 17**). On the other hand, non-SSRP1 target regions exhibited an equal distribution both in convergent and divergent Pol II (H3K4me3 High- Non-SSRP1 targets – **Fig. 17**). Upon knock-down of FACT, SSRP1 targets display higher levels of convergent transcription, and even more strikingly, a strong increase ($P < 10^{-27}$ and $P < 10^{-20}$) in divergent transcription compared to the Non-SSRP1 targets. To understand how FACT inhibits divergent transcription, we examined nucleosomal distribution over these promoters. Both high and low H3K4me3 genes, targeted by FACT, exhibited significantly increased levels of nucleosome occupancy ($P < 10^{-18}$ and $P < 10^{-4}$) upstream of the TSS, as opposed to the non-bound counterparts. Upon FACT depletion, nucleosome occupancy is decreased upstream of the TSS ($P < 10^{-8}$, $P < 10^{-6}$), resulting in a more open chromatin conformation and bi-directional travelling of Pol II.

Moreover, this increase in antisense transcription and loss of nucleosome occupancy is observed only in genes that are up-regulated upon FACT depletion (**Fig. 16 & Fig. 17 a,b**). These observations suggest that SSRP1-bound promoters construct directionality and inhibit divergent transcription by providing

a high level nucleosome barrier at the -1 and -2 nucleosomes. Such a scenario would be in agreement with data generated in *S. cerevisiae* where FACT regulates local nucleosomal stability to maintain the repression of non-coding transcripts(Feng et al., 2016). Taken together, FACT regulates unidirectional passage of RNA Pol II via nucleosome positioning.

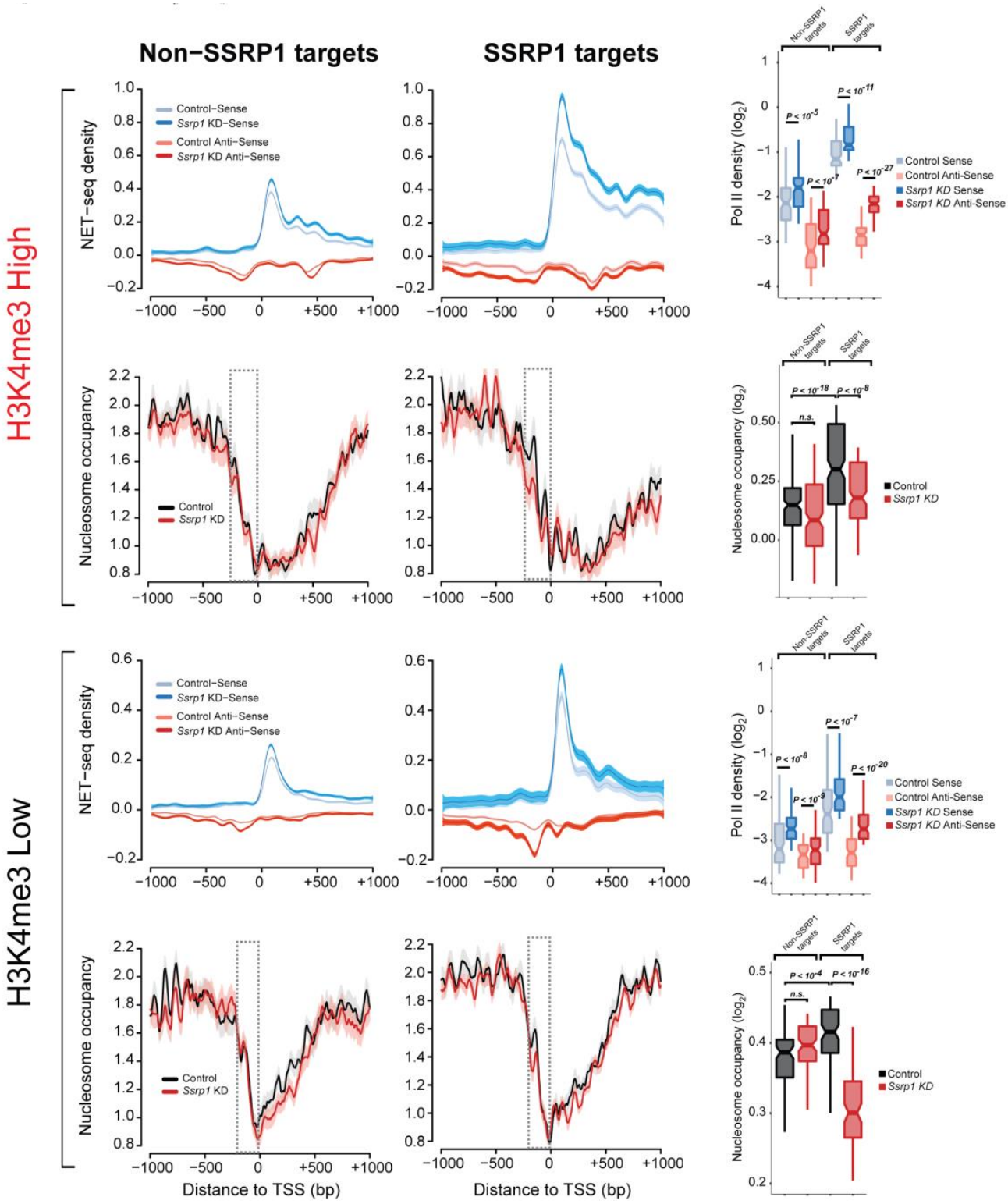
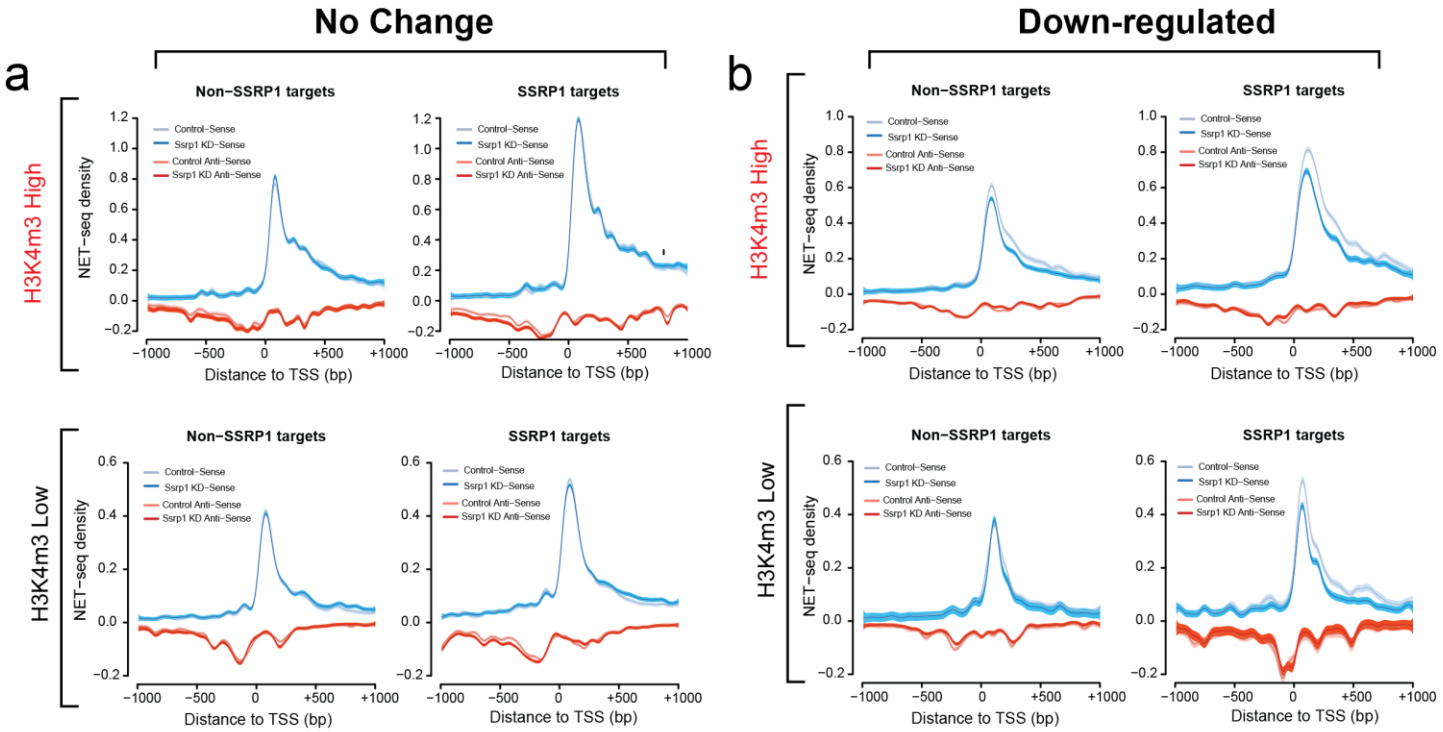


Figure 16**: FACT restricts divergent transcription via nucleosome deposition in the NFR.** NET-seq and nucleosome occupancy plots (Control and *Ssrp1* KD group) split by chromatin and FACT-bound status. Solid lines on the MNase-seq and NET-seq metaplots indicate the mean values, whereas the shading represents either the SE of the mean (MNase-seq) or the 95% confidence interval (NET-seq). Significant changes in Pol II/ nucleosome density were calculated using a Wilcoxon rank test.


No Change
Down-regulated
a
b
H3K4m3 High
H3K4m3 Low
Non−SSRP1 targets
SSRP1 targets
NET−seq density
Control−Sense
Ssrp1 KD−Sense
Control Anti−Sense
Ssrp1 KD Anti−Sense
Distance to TSS (bp)

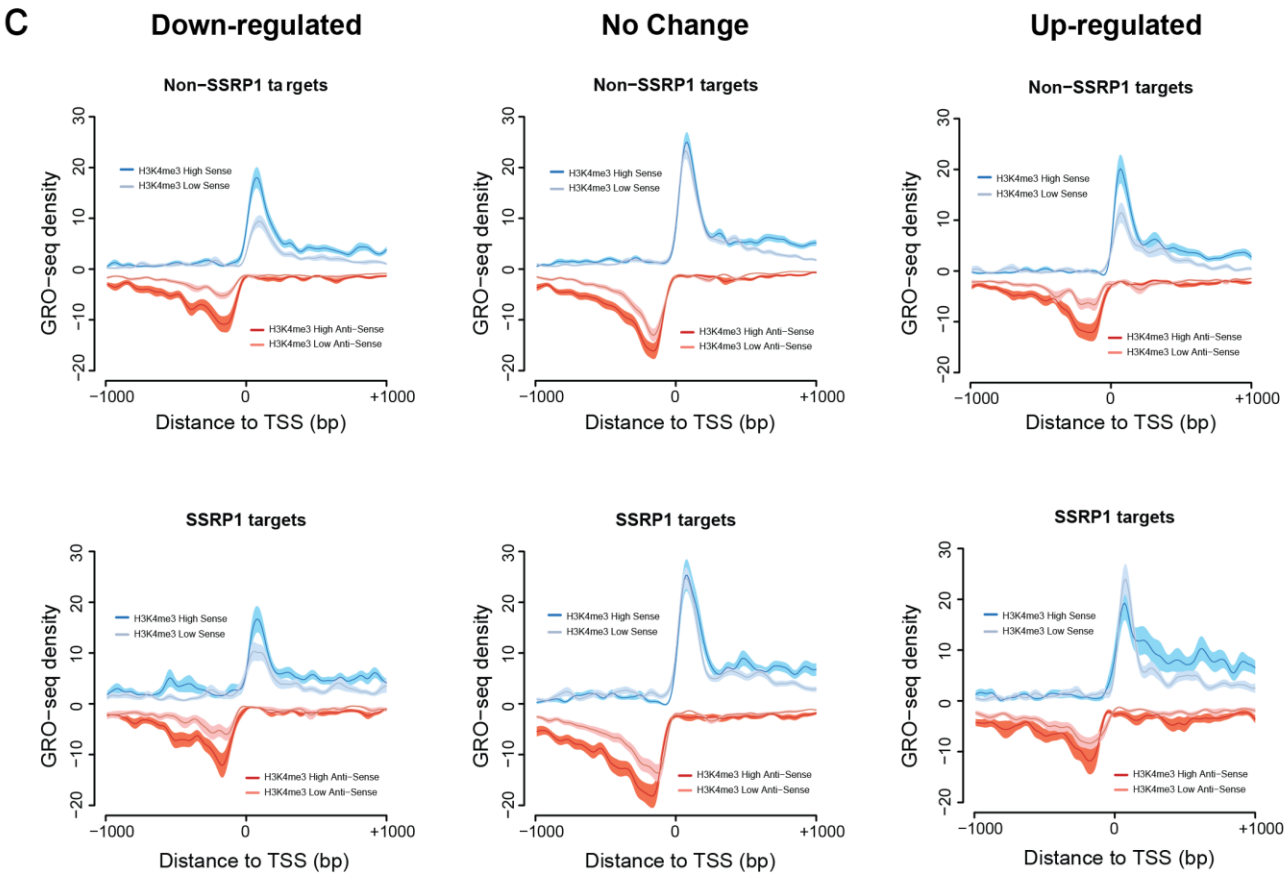

C
Down-regulated
No Change
Up-regulated
Non−SSRP1 targets
SSRP1 targets
GRO−seq density
H3K4me3 High Sense
H3K4me3 Low Sense
H3K4me3 High Anti−Sense
H3K4me3 Low Anti−Sense
Distance to TSS (bp)

Figure 17**: FACT enables broad distribution of RNA pol II over high H3K4me3 promoters.** **a**, NET-seq density plots for the “No Change” class split by FACT binding status (SSRP1 and Non-Ssrp1 targets) and condition (Control and *Ssrp1* KD). Occupancy was calculated relative to the TSS ± 2000 bp. **b**, Same as **a** but for the “Down-regulated” class. **c**, GRO-seq density plots derived from WT mESCs for the three different gene classes split by FACT binding status (SSRP1 and Non-Ssrp1 targets). Solid lines on the NET-seq and GRO-seq metaplots indicate the mean values, whereas the shading represents the 95% confidence interval.

Next, we focused on a single gene level to confirm whether nucleosome occupancy in the NFR hinders divergent transcription over FACT-bound promoters. Indeed, NET-seq antisense density seems to transit from convergent to divergent following depletion of *Ssrp1* mRNA levels, concomitantly with nucleosomal loss in the NFR (*Oct4* – SSRP1 target). On the contrary, *Psmb6* (Non-SSRP1 target) exhibits a slight increase of sense and anti-sense transcription in the *Ssrp1* KD state without a significant change in nucleosome density. The above data suggest that FACT regulates unidirectional passage of RNA Pol II via nucleosome positioning (**Fig. 18**).

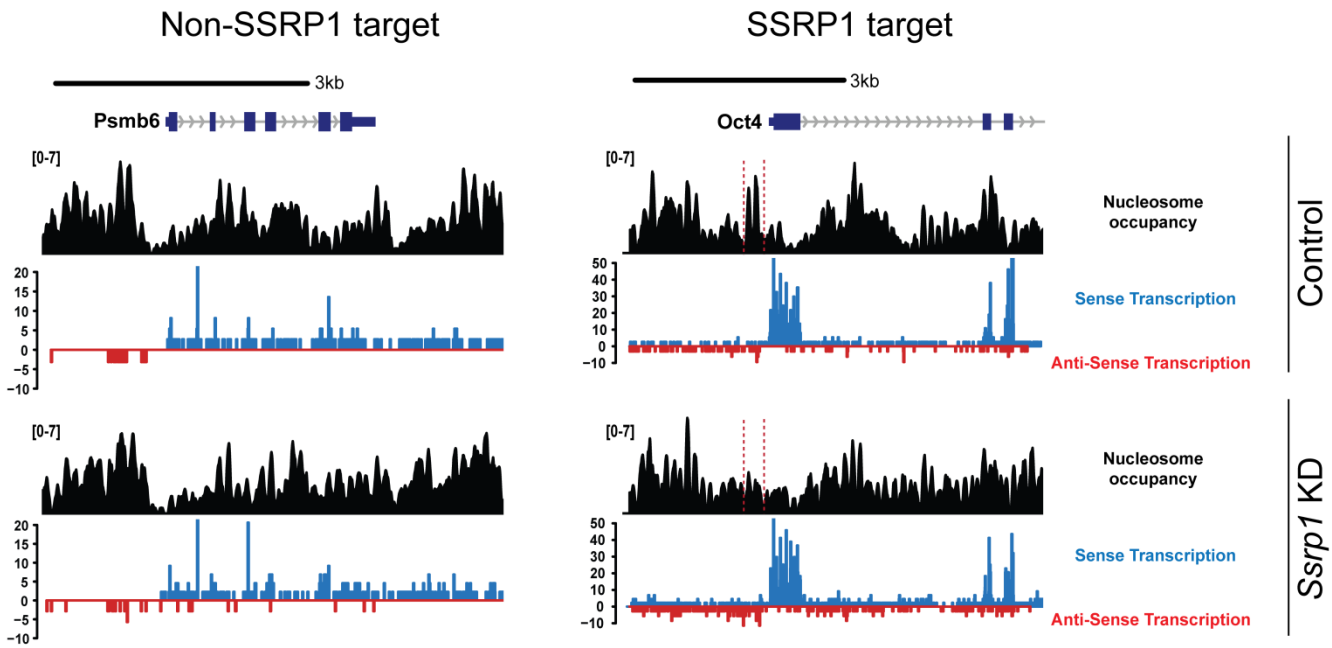


Figure 18**: Nucleosomal loss at the NFR instigates divergent transcription**. Transcriptional activity (NET-seq) and nucleosome occupancy (MNase-seq) of SSRP1 (*Oct4*) and non-SSRP1 (*Psmb6*) targets between Control and KD group. Nucleosomal loss in *Oct4* is indicated by a red-dotted line

4.9 FACT facilitates the alternative splicing of mRNA transcripts

High levels of H3K4me3 that span the first exon of the coding sequence have been implicated in efficient transcription (Bieberstein, Oesterreich, Straube, & Neugebauer, 2012). One potential explanation for this is that short exons are important for the precise recruitment of general transcription factors (GTFs). To further understand how FACT recruitment regulates gene expression, we examined the distribution of H3K4me3 and FACT over the first exon of all genes. Both H3K4me3 and FACT show high similarity in occupancy over the first exon, with the medium exon length class (> 250 and < 750 bp) displaying the highest occupancy over the first exon-intron junction (**Fig. 19 a,b**).

To determine the compendium of factors associating with FACT on actively transcribed regions, we performed an Immunoprecipitation (IP) for H3K4me3 followed by mass spectrometry between the Control and *Ssrp1* KD group. GTFs, cell cycle, and DNA proliferation components (MCM helicase complex) exhibited a significantly increased recruitment to H3K4me3 in the absence of FACT. Association of MCM and FACT has been previously reported in a cell cycle-dependent matter (Tan, Liu, Lin, & Lee, 2010), thus potentially explaining the increased cell proliferation we observed following depletion of FACT levels. Components of the spliceosome (e.g. CWC15, U2AF1) and several splicing factors (e.g. RBM17, SF3B6) displayed significantly decreased recruitment on H3K4me3 in the *Ssrp1* KD (**Fig. 19 c**). Interestingly, recruitment of components of the core spliceosome on H3K4me3 has been reported before in datasets using either peptides (Vermeulen et al., 2010) or modified nucleosomes (Bartke et al., 2010).

We thus speculated about a potential involvement of FACT in alternative splicing. In total, we identified 356 Exon skipping/inclusion and 97 Intronic retention events in the KD group (**Fig. 20 a,b,c**). Yeast *spt16* mutant strains have been shown before to retain intronic regions as a result of a compromised splicing machinery (Burckin et al., 2005). Skipped and retained exons were split according to exon length (< 500 bp and > 500 bp) and FACT-bound status (SSRP1 and Non-SSRP1 targets). Remarkably, the SSRP1-bound skipped exons exhibit lower length size in the alternative spliced "Long-exons" category (**Fig. 20 d**) compared to their non-bound counterparts (P = 0.02, Kolmogorov–Smirnov test), indicating an explicit facilitation of splicing for longer exons when FACT is present.

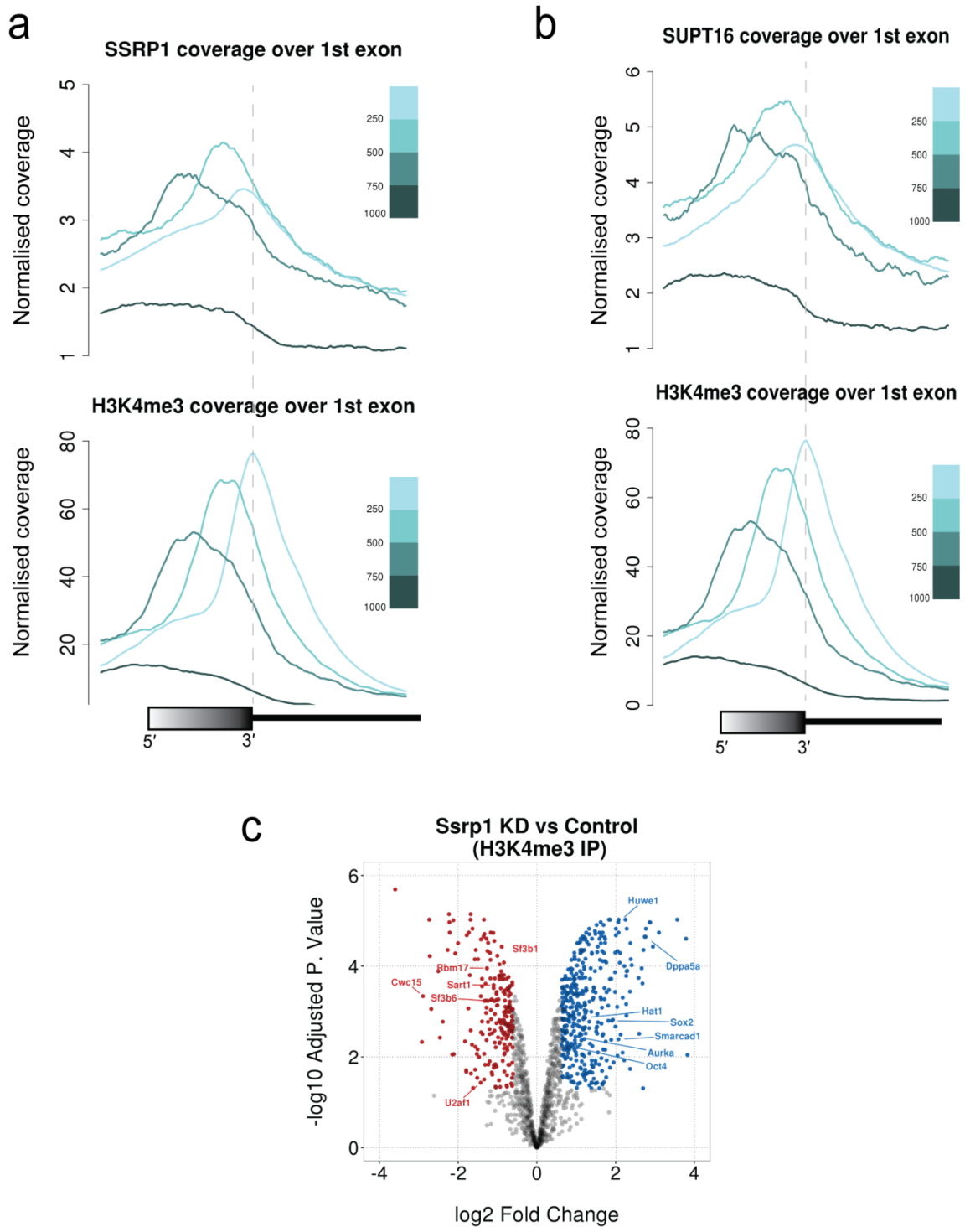


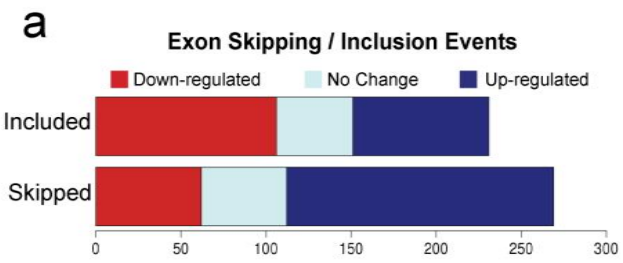
Figure 19: **FACT along with H3K4me3 serve as a protein docking site. a,** Average distribution of SSRP1 and H3K4me3 aligned to the 5' Splice Site (5'SS) of all genes grouped by first exon length. **b,** Same as **a** but for SUPT16, **c**, Volcano plot of depleted/enriched proteins at H3K4me3 following *Ssrp1* KD.

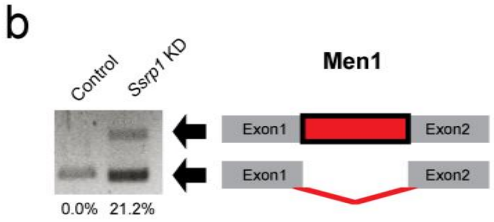
4.10 FACT is involved in the alternative splicing of gene isoforms

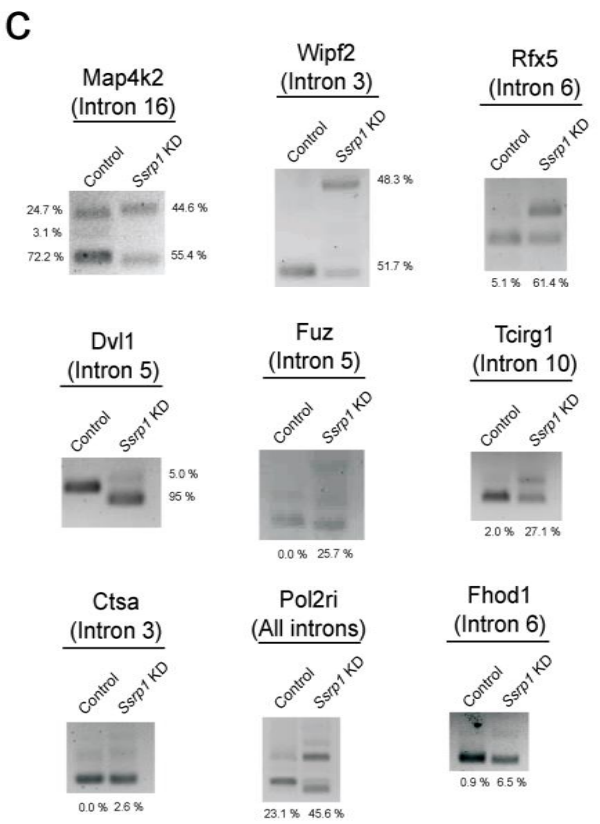
As differential exon-intron GC content and nucleosome occupancy are associated with exonic splicing (Amit et al., 2012; Kornblihtt et al., 2013; Sibley et al., 2016) ,we next sought to identify whether differential nucleosome occupancy between the control and the *Ssrp1* KD group is pivotal for alternative exon usage. We categorised the data according to the relevant exon skipping/ inclusion event (**Fig. 20 a**).

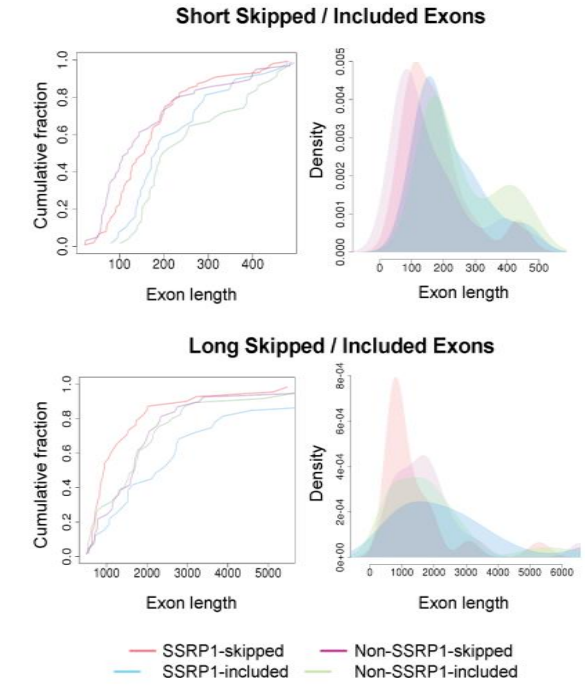
Altered nucleosome distribution can affect alternative exon usage (Iannone et al., 2015; Kornblihtt, Schor, Allo, & Blencowe, 2009). Hence, we interrogated nucleosomal distribution over 3' and 5' Splice Sites (SS). No significant difference was observed among the two conditions perhaps due to the presence of gene isoforms that encode different alternatively spliced exons. Nevertheless, we also determined RNA Pol II density derived from NET-seq data as this monitors the pausing degree at a single nucleotide resolution over splice sites (Mayer et al., 2015). Similarly to the MNase-seq data, no significant changes were observed among the two conditions due to low read density (**Fig. 21**).

Despite NET-seq being a powerful technique in determining the position of Pol II genome-wide at a single nucleotide resolution, the library preparation steps are characterised by a substantial loss of genomic material that causes the emergence of numerous PCR duplicates in the final product (Mayer & Churchman, 2016). Although, here we present an improved version of NET-seq, future work should focus on optimally adjusting the library preparation with regard to maximising Pol II coverage.


a
Exon Skipping / Inclusion Events
Down-regulated
No Change
Up-regulated
Included
Skipped
0
50
100
150
200
250
300


b
Men1
Control
Ssrp1 KD
Exon1
Exon2
Exon1
Exon2
0.0%
21.2%


c
Map4k2 (Intron 16)
Control
Ssrp1 KD
24.7 %
3.1 %
72.2 %
44.6 %
55.4 %
Wipf2 (Intron 3)
48.3 %
51.7 %
Rfx5 (Intron 6)
5.1 %
61.4 %
Dvl1 (Intron 5)
5.0 %
95 %
Fuz (Intron 5)
0.0 %
25.7 %
Tcirg1 (Intron 10)
2.0 %
27.1 %
Ctsa (Intron 3)
0.0 %
2.6 %
Pol2ri (All introns)
23.1 %
45.6 %
Fhod1 (Intron 6)
0.9 %
6.5 %


d
Short Skipped / Included Exons
Cumulative fraction
Exon length
Density
Exon length
Long Skipped / Included Exons
Cumulative fraction
Exon length
Density
Exon length
SSRP1-skipped
SSRP1-included
Non-SSRP1-skipped
Non-SSRP1-included

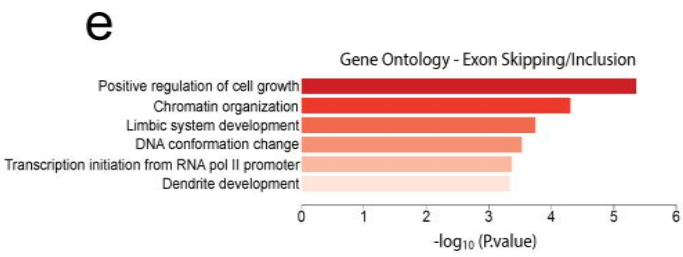

e
Gene Ontology - Exon Skipping/Inclusion
Positive regulation of cell growth
Chromatin organization
Limbic system development
DNA conformation change
Transcription initiation from RNA pol II promoter
Dendrite development
0
1
2
3
4
5
6
-log10 (P.value)

Figure 20: **FACT facilitates alternative splicing of RNA transcripts. a,** Average distribution of SSRP1, SUPT16, and H3K4me3 aligned to the 5' Splice Site (5'SS) of all genes grouped by first exon length. **b,** Volcano plot of depleted/enriched proteins at H3K4me3 following *Ssrp1* KD. **c**, Barplots representing the number of included/skipped exons categorized by their gene expression status (red: “Down-regulated”, cyan: “No change”, blue: “Up-regulated”). In total, we have identified 149 included and 207 skipped exon events in the *Ssrp1* KD group. **d**, Graphical representation of an intronic retention event (*Men1*) in the KD group. Also, analysis of intron inclusion events or isoform switches after FACT depletion. Unspliced transcript percentage was measured according to band intensity. **e,** Gene ontology analysis of transcripts (FACT-bound) that display alternative exon usage between the two conditions.

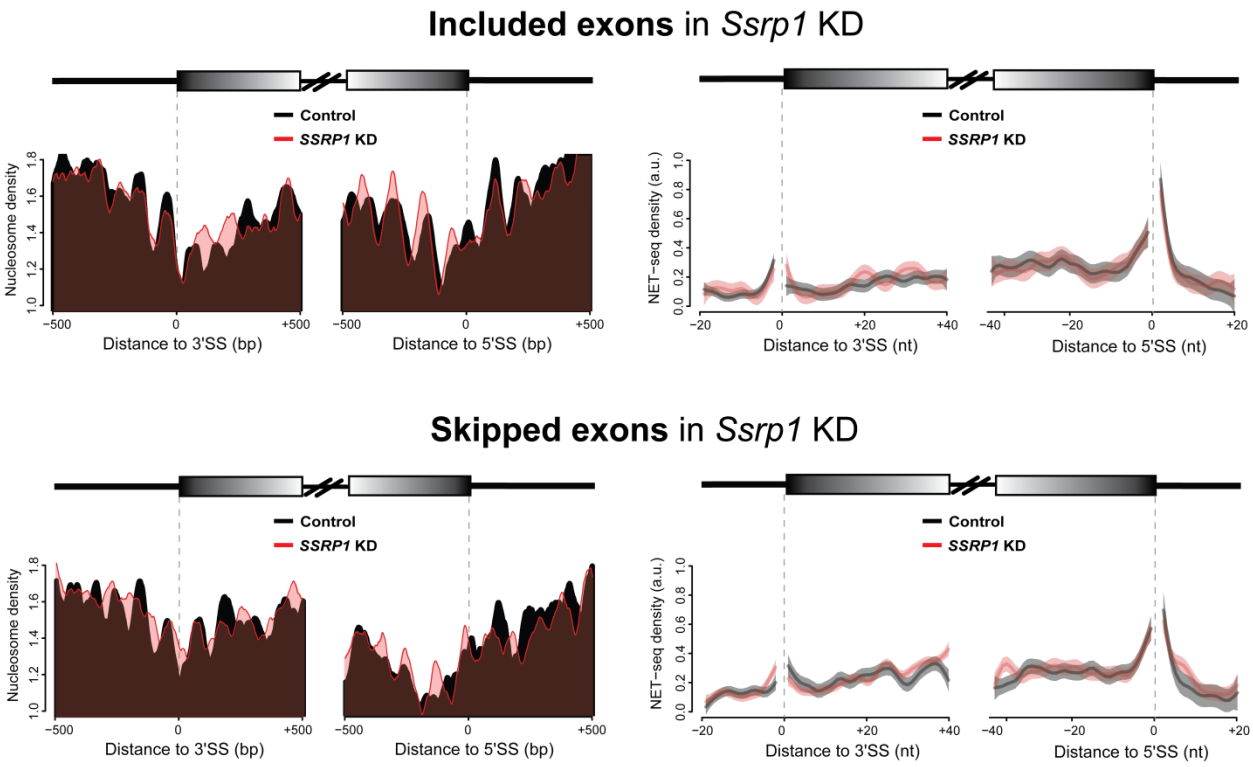


Figure 21: **Nucleosomal occupancy along with Pol II density drive alternative exon usage.** Nucleosome occupancy and NET-seq density of FACT-bound exons that are differentially spliced upon depletion of *Ssrp1***.** Solid lines on the NET-seq meta-exon plots indicate the mean values, whereas the shading represents the 95% confidence interval.

4.11 ES cells differentiate more efficiently into the neuronal lineage upon FACT depletion

A fraction of the differential gene isoforms generated in the KD group is over-represented in limbic system and dendrite development pathways (Fig. 20e). To confirm this, we induced ES cell differentiation towards a neuronal lineage via embryoid body formation and treatment with retinoic acid (Bibel, Richter, Lacroix, & Barde, 2007). We created early stage Neural Precursor Cells (NPCs – 3 days into the differentiation process) that express key neurogenesis markers (Pax6, Nes, Tubb3) but still maintain FACT and key pluripotency factors at a high level (**Fig. 22a**). A quarter of the up-regulated genes in ES cells after *Ssrp1* KD overlaps with the up-regulated genes instigated by neuronal differentiation ($P < 10^{-13}$, Fisher's exact test; **Fig. 22b**) and are over-represented in pathways involved in neuronal development. We then split these over-lapping genes according to FACT-bound status and interrogated nucleosome/ Pol II occupancy. Interestingly, all the neurogenesis-associated genes (based on gene ontology enrichment) found on this list are bound by FACT (n=38). Intriguingly, only these genes bound by FACT show a significant increase ($P < 10^{-15}$) in divergent anti-sense transcription concomitantly with loss of a nucleosomal impediment ($P < 10^{-46}$) (**Fig. 22c**). Up-regulation of FACT-bound genes in its absence coincides with a loss of nucleosome upstream of the TSS and an increase in antisense transcription. To test if de-repression of silenced genes in the absence of FACT follows a similar mechanism, we analysed nucleosome occupancy and RNA Pol II profiles over these genes. The observed changes in nucleosome and RNA Pol II occupancy were even more pronounced in this gene class, and exhibited similar patterns (**Fig. 23**). Divergent transcription has been suggested to enhance

transcription by opening up promoters and allowing binding of multiple transcription factors (Scruggs et al., 2015).

Finally, to understand the biological impact of FACT depletion on neuronal development, we depleted *Ssrp1* levels at the onset of neuronal differentiation and performed immunofluorescence for neurogenesis (β3-Tubulin) and dendritic (MAP2) markers at the same time point as the RNA-seq experiment. *SSRP1 KD* caused a substantial increase in the expression of those markers, indicating that loss of FACT function primes ES cells for the neuronal lineage and enhances early neuronal differentiation (**Fig. 22d**).

Based on a genome-wide RNAi screen in Drosophila, Ssrp1 was discovered as a gene that will lead to hyper-proliferation of neuroblasts (Neumüller et al., 2011) – a neuronally committed cell type. Usually, neuroblasts divide asymmetrically at every stage, creating one cell that continues being a neuroblast, and one cell that becomes the Ganglion Mother Cell, which goes on to divide into neurons and glia cells (Brand & Livesey, 2011). Depletion of FACT shifts this balance and leads to more neuroblasts in expense of neurons. A comparison of expression patterns in the early developing mouse brain identified a set of only 13 genes, including *Ssrp1* with high correlation of expression in the proliferating cells of the VZ of the neocortex at early stages of development (Vied et al., 2014). This is a transient embryonic layer of tissue containing neural stem cells (Rakic, 2009) and a place for neurogenesis during development dependent on the Notch pathway (Rash, Lim, Breunig, & Vaccarino, 2011). Our data demonstrates that FACT-depleted ES cells differentiate much more efficiently into early neuronal precursors. Taken together, the data suggest an important role for FACT activity during neuronal differentiation and the proper levels of FACT might assist in

balancing proliferation speed and timing of differentiation processes. Another transcriptional elongation complex that interacts with FACT and was shown to lead to hyper-proliferation of Drosophila neuroblasts is the Paf1 complex (Neumüller et al., 2011). Interestingly, depletion of the Paf1 complex in mouse ES cells leads to a different transcriptional outcome than FACT depletion. Paf1 down-regulation leads to loss of pluripotency, in part through down-regulation of the key pluripotency factors, including *Nanog*, *Oct4*, and *Sox2* (Ding et al., 2009). Like FACT, the Paf1 complex can work as an activator and repressor of gene transcription in ES cells, leading to deregulation of around 1200 genes. This comparison between the FACT and Paf1-complex suggest that transcriptional regulators that impact the ES cell transcriptome in different ways can have similar roles during differentiation processes.

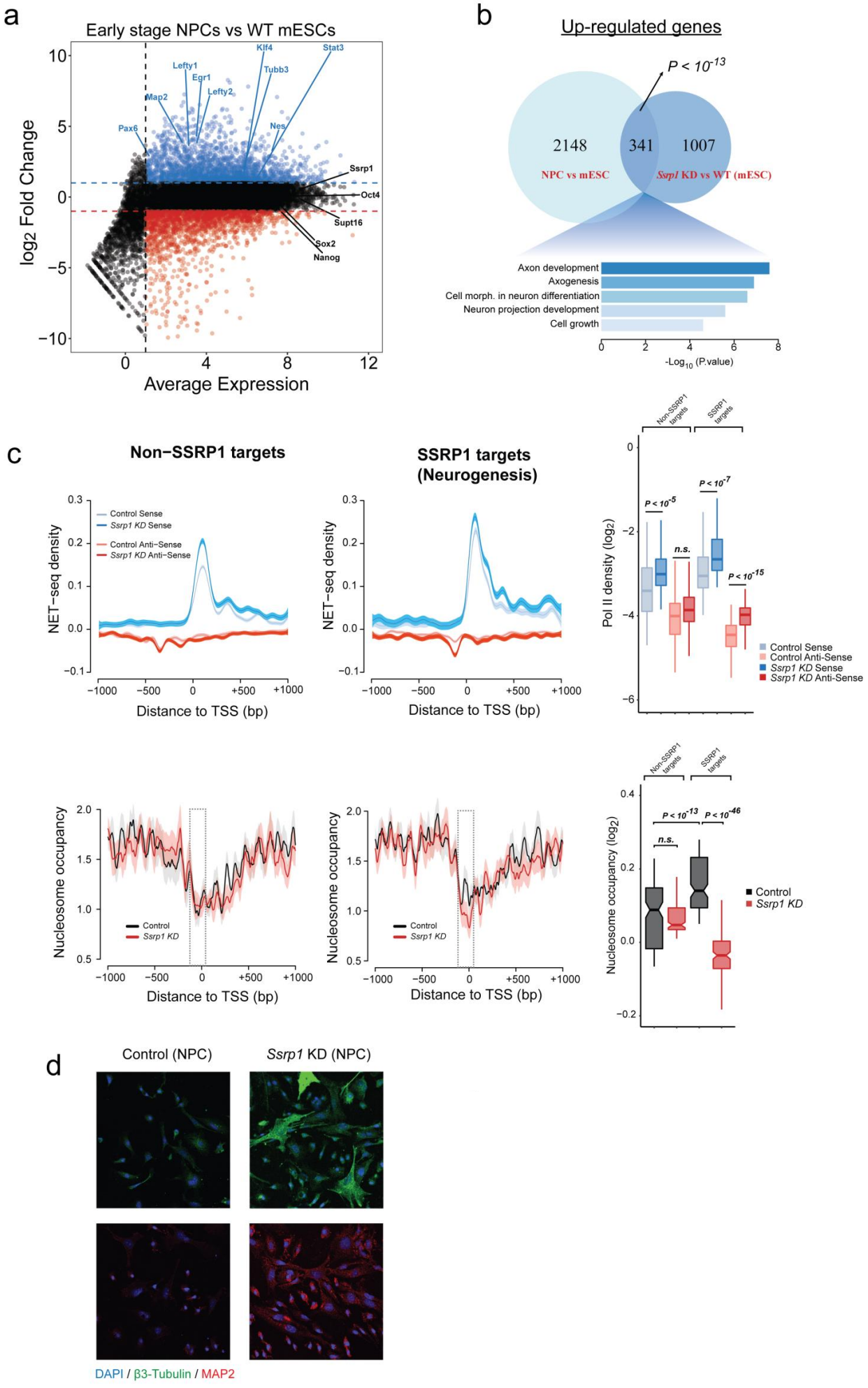

a
Early stage NPCs vs WT mESCs
log2 Fold Change
Average Expression
Klf4
Stat3
Lefty1
Tubb3
Egr1
Lefty2
Map2
Pax6
Nes
Ssrp1
Oct4
Supt16
Sox2
Nanog
b
Up-regulated genes
P < 10-13
2148
341
1007
NPC vs mESC
Ssrp1 KD vs WT (mESC)
Axon development
Axogenesis
Cell morph. in neuron differentiation
Neuron projection development
Cell growth
-Log10 (P.value)
c
Non−SSRP1 targets
SSRP1 targets (Neurogenesis)
Control Sense
Ssrp1 KD Sense
Control Anti−Sense
Ssrp1 KD Anti−Sense
NET−seq density
Distance to TSS (bp)
Non-SSRP1 targets
SSRP1 targets
P < 10-5
n.s.
P < 10-7
P < 10-15
Pol II density (log2)
Control
Ssrp1 KD
Nucleosome occupancy
Nucleosome occupancy (log2)
P < 10-13
P < 10-46
d
Control (NPC)
Ssrp1 KD (NPC)
DAPI / β3-Tubulin / MAP2

Figure 22: **FACT regulates neurogenesis through Pol II / nucleosome dynamics. a**, MA plot following depicting differential expression in NPCs versus WT ES cells. Up-regulated genes are highlighted in blue whereas down-regulated genes are highlighted in red. **b**, Venn diagram showing the overlap of up-regulated genes between NPC vs mESCs and Control vs *Ssrp1* KD mESCs. **c**, NET-seq and nucleosome occupancy plots (Control and *Ssrp1* KD group) split by chromatin and FACT-bound status. Solid lines on the MNase-seq and NET-seq metaplots indicate the mean values, whereas the shading represents either the SE of the mean (MNase-seq) or the 95% confidence interval (NET-seq). Significant changes in Pol II/ nucleosome density were calculated using a Wilcoxon rank test. **d**, Immunofluorescence (IF) analysis of early stage NPCs following *Ssrp1* depletion. (Blue) DAPI, nuclei; (Green) β3-Tubulin, neurons; (Red) MAP2, dendrites.

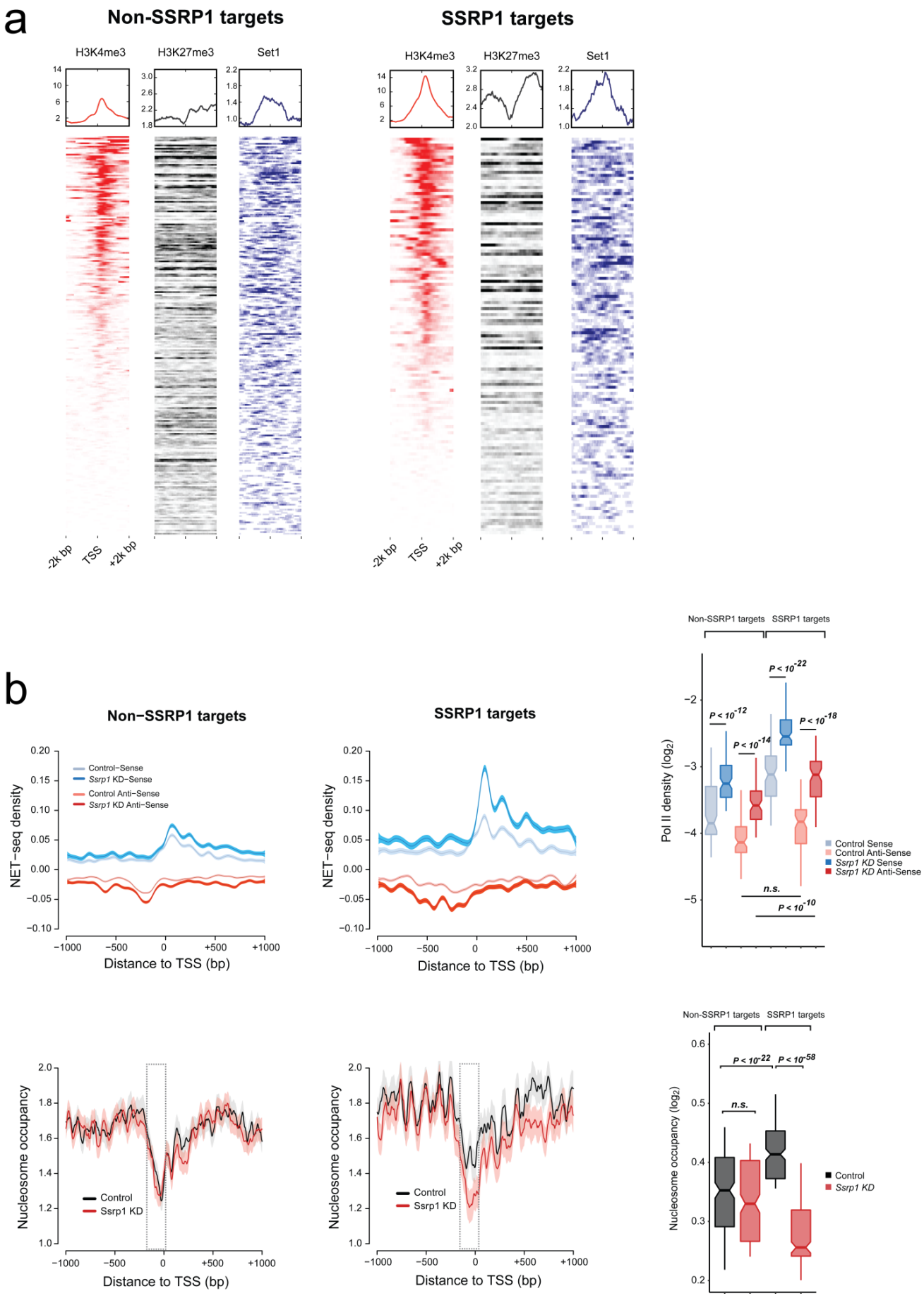

a
Non-SSRP1 targets
SSRP1 targets
H3K4me3
H3K27me3
Set1
-2k bp
TSS
+2k bp
b
Non−SSRP1 targets
SSRP1 targets
NET−seq density
Control−Sense
Ssrp1 KD−Sense
Control Anti−Sense
Ssrp1 KD Anti−Sense
Distance to TSS (bp)
Pol II density (log2)
Control Sense
Control Anti-Sense
Ssrp1 KD Sense
Ssrp1 KD Anti-Sense
n.s.
Nucleosome occupancy
Control
Ssrp1 KD
Nucleosome occupancy (log2)
Control
Ssrp1 KD

Figure 23**: Distribution of chromatin marks, nucleosomes, and Pol II over lowly expressed/ repressed genes. a**, Occupancy heatmaps and profiles (**Figure 11b**– Class 3) for H3K4me3, H3K27me3, and Set1 split by chromatin (High H3K4me3 and Low H3K4me3) and FACT binding status (SSRP1 and Non-Ssrp1 targets). Occupancy was calculated relative to the TSS ± 2000 bp. **b**, NET-seq and nucleosome occupancy plots (Control and *Ssrp1* KD group) split by chromatin and FACT-bound status. Solid lines on the MNase-seq and NET-seq metaplots indicate the mean values, whereas the shading represents either the SE of the mean (MNase-seq) or the 95% confidence interval (NET-seq). Significant changes in Pol II/ nucleosome density were calculated using a Wilcoxon rank test.

5 Conclusion and Outlook

As described above, the presented study has contributed in elucidating the molecular mechanisms under which FACT maintains pluripotency and governing of RNA Pol II *in vivo*. We have applied a wide range of sequencing and proteomics experiments (ChIP-seq, RNA-seq, MNase-seq, NET-seq, and Mass spectrometry) to fully characterise the interplay between FACT, nucleosomes, RNA Pol II, and chromatin dynamics. Our main findings are;

- FACT is involved in transcriptional elongation and is tightly linked to active gene expression. In addition, we have characterised FACT as both an activator and repressor of transcription.
- FACT ablation leads to little changes in nucleosome occupancy genome-wide. In fact, it seems to regulate expression of a specific class of genes involved in embryonic development (e.g. Yamanaka factors) by placing nucleosomes upstream of the TSS and hence maintaining a closed chromatin conformation state.
- FACT represses specifically genes involved in neurogenesis. Its depletion primes neural precursor formation by activation and alternative splicing of the above genes.
- This nucleosome deposition upstream of the TSS hinders divergent travelling of Pol II.
- FACT and several nucleosome remodellers control gene expression whereas the occupancy degree of the later ones seems to predict the change in gene expression levels in the absence of FACT.

Taken together, FACT maintains pluripotency through nucleosome deposition and Pol II governing.

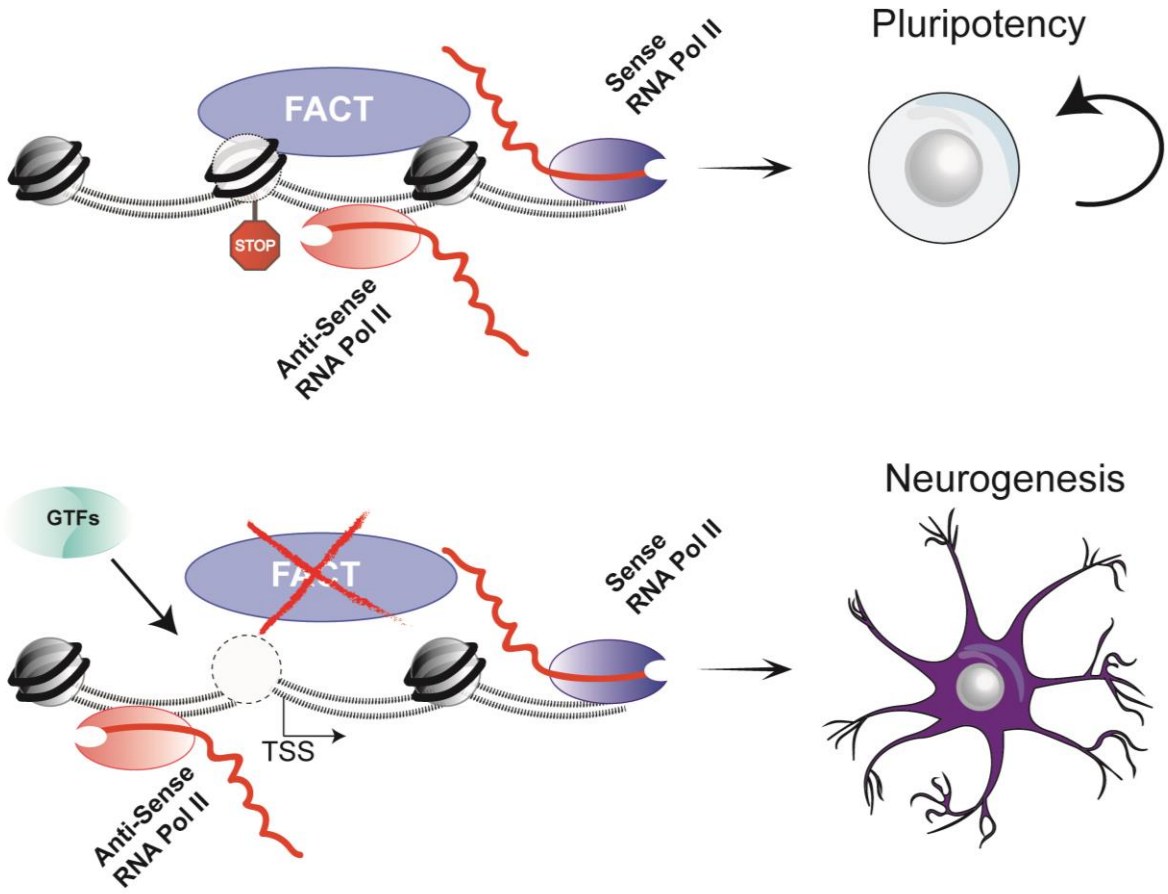


Figure 24: **Model of the suppressive role of FACT in gene expression and maintenance of pluripotency**. Upper; FACT places a nucleosomal barrier at the promoter region of genes involved in embryogenesis/ neurogenesis that hinders divergent travelling of Pol II and enables a closed chromatin conformation state. Lower; In the absence of FACT, the nucleosomal barrier is alleviated, thus allowing bi-directional travelling of Pol II, recruitment of GTFs, increased gene expression, and ultimately, activation of neurogenesis cues.

We have confirmed FACT's involvement in transcriptional elongation. However, we have discovered that FACT is not contributing to the release of initiating Pol II to a successive elongation like NELF and P-TEFb. Since FACT binds also upstream of the TSS, could it also be involved in transcriptional initiation and assembly of the pre-initiation complex (PIC)?

To address this question we would require interrogation of FACT occupancy and of other initiation complexes using more sophisticated/high resolution mapping approaches (ChIP-exo/ ChIP-nexus - (Shao & Zeitlinger, 2017)) followed by chemical inhibition of transcriptional elongation.

Moreover, recent reports have shown that Pol II variants display different binding on nascent RNA as opposed to genomic DNA (Nojima et al., 2015). Could this be the same for FACT as well as for other elongation factors? Is there perhaps an intrinsic function for FACT that allows dynamic phosphorylation of Pol II as it synthesizes newly formed RNA? More specific and directed techniques would be required to be developed to structurally interrogate co-binding of FACT and Pol II during nascent RNA synthesis.

6 Materials

Growth media and antibiotics

DMEM + GlutaMax	Gibco
Fetal Bovine Serum (FBS)	Gibco
MEM non- essential amino acids	Gibco
Penicillin/ Streptomycin	Gibco
2-mercaptoethanol	Gibco
Leukaemia Inhibitory Factor (LIF)	eBioscience
Retinoic Acid (RA)	Sigma
Puromycin	Sigma

Buffers

ChIP-seq

Cell Lysis Buffer
5 mM Tris
pH8.0, 85 mM KCl
0.5% NP40

Nuclei Lysis Buffer
1% SDS
10 mM EDTA
50 mM Tris HCl

IP buffer
0.01% SDS
1.1% Triton-X-100
1.2 mM EDTA
16.7 mM Tris HCl,
67 mM NaCl

TSE -150
1% Triton-X-100
0.1% SDS
2 mM EDTA
20 mM Tris HCl
150 mM NaCl

TSE -500
1% Triton-X-100
0.1% SDS
2 mM EDTA

		20 mM Tris HCl
		500 mM NaCl
	LiCl buffer	0.25 M LiCl
		1% NP-40
		1% dioxycholate
		1mM EDTA
		10 mM Tris HCl
	Elution Buffer	1% SDS
		0.1 M $NaHCO_3$
NET-seq	Cytoplasmic lysis buffer	0.15% (vol/vol) NP-40
		10 mM Tris-HCl (pH 7.0)
		150 mM NaCl
		1× protease inhibitor mix (50×)
		25 µM α-amanitin
		10 U SUPERase
		RNase-free H2O
	Sucrose buffer	50% (wt/vol) filter-sterilized sucrose
		10 mM Tris-HCl (pH 7.0)
		150 mM NaCl
		1× protease inhibitor mix (50×)
		25 µM α-amanitin
		20 U SUPERase
		RNase-free H2O
	Nuclei wash buffer	1 mM EDTA
		0.1% (vol/vol) Triton X-100
		1× protease inhibitor mix (50×)
		25 µM α-amanitin
		40 U SUPERase
		PBS
	Glycerol buffer	20 mM Tris-HCl (pH 8.0)
		75 mM NaCl
		0.5 mM EDTA
		100% (vol/vol) filter-sterilized glycerol
		0.85 mM filter sterilized DTT
		1× protease inhibitor mix (50×)
		25 µM α-amanitin
		10 U SUPERase
		RNase-free H2O

Nuclei lysis buffer	1% (vol/vol) NP-40
	20 mM HEPES (pH 7.5)
	0.1 M EDTA
	300 mM NaCl
	10 M filter sterilized urea
	1 mM filter-sterilized DTT
	1× protease inhibitor mix (50×)
	25 µM α-amanitin
	10 U SUPERase
	RNase-free H2O
Chromatin resuspension solution	1× protease inhibitor mix (50×)
	25 µM α-amanitin
	20 U SUPERase
	PBS
Bind / wash buffer	5 mM Tris-HCl (pH 7.0),
	2 M NaCl
	1 mM EDTA
	0.2% (vol/vol) Triton X-100
	RNase-free H2O

NET-seq reagents

- α-Amanitin (Sigma)
- TBE buffer, 10× (Life Technologies, cat. no. 15581-044)
- Urea, (Sigma-Aldrich, cat. no. U6504)
- Sodium carbonate anhydrous, proteomics grade (VWR, cat. no. M138)
- Sodium bicarbonate (VWR, cat. no. 3509)
- Sodium acetate, RNase-free (3 M; Life Technologies, cat. no. AM9740)
- NaCl, RNase-free (5 M; Life Technologies, cat. no. AM9760G)
- EDTA, RNase-free (0.5 M; Life Technologies, cat. no. AM9260G)
- Tris-HCl, RNase-free (1 M, pH 7.0; Life Technologies, cat. no. AM9850G)
- Tris-HCl, RNase-free (1 M, pH 8.0; Life Technologies, cat. no. AM9855G)
- NaOH solution (1.0 N; Sigma-Aldrich, cat. no. S2770)
- HCl, hydrochloric acid concentrate (Sigma-Aldrich, cat. no. 38282)
- SUPERase.In (Life Technologies, AM2696)
- Protease inhibitor mix cOmplete, EDTA-free (Roche, 11873580001)
- Benzonase nuclease (Sigma-Aldrich, cat. no. E1014)

- miRNeasy mini kit (50; Qiagen, cat. no. 217004)
- T4 RNA ligase 2, truncated (NEB, cat. no. M0242S)
- SYBR Gold nucleic acid gel stain (10,000× concentrate; Life Technologies, cat. no. S-11494)
- RNA control ladder (0.1–2 kb; Life Technologies, cat. no. 15623-100)

- DNA control ladder (10 bp; Life Technologies, cat. no. 10821-015)
- TBE-urea (TBU) denaturing sample buffer, 2× (Life Technologies, cat. no. LC6876)
- SuperScript III first-strand synthesis system (Life Technologies, cat. no. 18080-051)
- CircLigase ssDNA ligase; (Biozym, cat. no. CL4111K)
- Saline–sodium citrate (SSC), 20× (Life Technologies, cat. no. AM9763)
- Phusion high-fidelity (HF) DNA polymerase (2,000 U/ml; NEB, cat. no. M0530S)

- TBE-urea gels, 15% (wt/vol) (Life Technologies, cat. no. EC68852BOX)
- TBE-urea gels, 10% (wt/vol) (Life Technologies, cat. no. EC68752BOX)
- TBE gels, 8% (wt/vol) (Life Technologies, cat. no. EC62152BOX)
- Mini-Cell polyacrylamide gel box, XCell SureLock (Life Technologies, cat. no. EI0001)
- Black gel box (LI-COR, cat. no. 929-97301)

NET-seq depletion oligos

Symbol	Transcript	DNA sequence (5'-3')
Snord49a	snoRNA	AGTCAGCCAGGAGCAGTTATCGTCAGTTATCGAC
Rn45s	rRNA	GAGAGCCGCCCGAACGACCGACTTCCCTACGGGCCC
Snord65	snoRNA	CTTCAGAAAACCATAGGCTCACCACTACCAATCT
Snord82	snoRNA	GAACCATGGGGTTGAAATGAAATATGCTGATGTGCT
Snord49b	snoRNA	GTCAGCTAACTAGGGATGTCGTCAGTTGTCGCAT
Snord2	snoRNA	AGTGATCAGCAAGAGTATTCTCTTCATTTCAGGTCA
Snord99	snoRNA	TCTCAGTCCCATATCCGCATTTCTCATCCATAGA
Snord95	snoRNA	CAGCTCAGAAACAGCCTCTGGATTTCAGCAAAGCAA
Snord55	snoRNA	CGTGGGGAAGCCAACCTTGGAGAGCTGAGCGTGC
Snord68	snoRNA	CATCAGATGGAAAAGGGTTCAAAAGTACTTTCAT
Snord32a	snoRNA	GACTGTGAGATCAACCCATGCACCGCTCTGAGACTC
Snord87	snoRNA	GTTTCTTTGAAGAGAGAATCTTAAAAGACTGAGA
Rmrp	ncRNA	CGCACCAACCACACGGGGCTCATTCTCAGCGCGGCTAC
Snord100	snoRNA	CTCGCTGAGGAAACTGCACGTCACCCTCCTGAAA
Snora68	snoRNA	GTGCAGTGCCCCCCAGAGTGAATCAGTAGGCTCTACAGAA
Rnu3a	snRNA	AACCACTCAGACTGTGTCCTCTCCCTCTCAACCCTCAA
Snord42b	snoRNA	GAGACCTGTGATGTCTTCAAAGGAACCACTGATG
Snord83b	snoRNA	TGAGGAATTATTCCCTGTTGCCTTCCTTCTGAGA
Snord110	snoRNA	TTGCTCAGACACATGGAGTCGTCAGTGATCTCTCAGGG
Snord47	snoRNA	CCTCAGAAATAAAATGGAACGGTTTAAAGGTGAT

Antibodies

SSRP1	Biolenends
SUPT16	Cell signalling
Alpha-tubulin	Abcam
β3-Tubulin	Cell signalling
MAP2	Millipore
H3K4me3	Active motif

7 Methods

Cell culture. The E14 cell line (mESCs) was cultured at 37 °C, 7.5% CO_2, on 0.1% gelatin coated plates, in DMEM + GlutaMax™ (Gibco) with 15% fetal bovine serum (Gibco), MEM non- essential amino acids (Gibco), penicillin/streptomycin (Gibco), 550 µM 2-mercaptoethanol (Gibco), and 10 ng/ml of leukaemia inhibitory factor (LIF) (eBioscience). HEK293T, N2a, MEFs, NIH3T3, and B16 cell lines were cultured at 37 °C, 5% CO_2 in DMEM + GlutaMax™ (Gibco) with 10% fetal bovine serum (Gibco), and penicillin/streptomycin (Gibco). Early Neuronal Precursor Cells (NPCs) were generated as previously described (Bibel et al., 2007). Briefly, embryoid bodies were created with the hanging drop technique and were further treated with 1 µM retinoic acid (RA) for 4 days. RA-treated embryoid bodies were trypsinised and cultured in DMEM + GlutaMax™ (Gibco) with 15% fetal bovine serum without LIF for 3 days.

Depletion of SSRP1 from mESCs via shRNA and RNA preparation. E14 were transfected with lentiviral vectors containing either a scramble Control or *Ssrp1* shRNAs (MISSION® shRNA, Sigma) with the following sequences:

Scramble Control	CCGGGCGCGATAGCGCTAATAATTTCTCGAGAAATTATTAGCGCTATCGCGCTTTTT
shRNA 1 (*Ssrp1*)	CCGGGCGTACATGCTGTGGCTTAATCTCGAGATTAAGCCACAGCATGTACGCTTTTTG
shRNA 2 (*Ssrp1*)	CCGGGCAGAGGAGTTTGACAGCAATCTCGAGATTGCTGTCAAACTCCTCTGCTTTTTG
shRNA 3 (*Ssrp1*)	CCGGCCGTCAGGGTATCATCTTTAACTCGAGTTAAAGATGATACCCTGACGGTTTTTG
shRNA 4 (*Ssrp1*)	CCGGCCGTCAGGGTATCATCTTTAACTCGAGTTAAAGATGATACCCTGACGGTTTTTG

A combination of two different *Ssrp1* shRNAs was used (1&2, 3&4) at a time and depletion was quantified via western blotting using a monoclonal anti-Ssrp1 antibody (Biolegends). Anti-alpha Tubulin was used as a reference control. The 1&2 combination was used for subsequent experiments as it yielded higher depletion of SSRP1 levels (Extended Fig 2a,b). Generation and transfection of shRNA vectors was done as previously described(Ramezani & Hawley, 2002).

Fourty-eight hours (48h) after transfection, puromycin (2 µg /ml) selection was applied for an additional 24h period, before cell collection and RNA preparation. Total RNA was obtained via phenol-chloroform extraction (QIAzol Lysis Reagent – QIAGEN) followed by purification via Quick-RNA™ MicroPrep (Zymo Research). Library preparation and ribosomal depletion were performed via the NEBNext Directional RNA Ultra Kit (NEB) and the RiboZero Kit (Illumina)

according to the manufacturer's instructions, respectively. Four different biological replicates (Control or SSRP1-depleted mESCs) were prepared and processed for transcriptome analysis.

MTT proliferation assay. 48h after transfection, different cell densities ($3x10^4$, $2x10^4$, $1x10^4$) were seeded on 96-well plates (Sarstedt) along with puromycin (2 µg /ml). Twenty-four hours later, the CellTiter 96® Non-Radioactive Cell Proliferation Assay kit (Promega) was used according to the manufacturer's instructions in order to assess the rate of cell proliferation between the two conditions (Control, *Ssrp1* KD). Statistical analysis was performed using a two-tailed *t-test*.

Transcriptome analysis in SSRP1-depleted mESCs. Sequenced reads were aligned to the mm10 genome via STAR (v 2.4.1b)(Dobin et al., 2013). Gene and exon counts were obtained from featureCounts of the Rsubread package (R/Bioconductor). Only reads with CPM (counts per million) > 1 were kept for subsequent analysis. Counts were normalised using the internal TMM normalisation in edgeR(Robinson, McCarthy, & Smyth, 2009) and differential expression was performed using the limma(Ritchie et al., 2015) package. Significant genes with an absolute logFC > 1 and Adjusted P.Value < 0.01 were considered as differentially expressed. The "No Change" gene class (n=2,179) was obtained from genes with an Adjusted P.Value > 0.05. The diffSplice function implemented in limma was used to identify differentially spliced exons between the two conditions. Significant exons with an FDR < 0.001 were

considered as differentially spliced. Retention Introns were identified using the MISO (Katz, Wang, Airoldi, & Burge, 2010) (Mixture of Isoforms) probabilistic framework.

Retention intron events. We verified the presence of retained introns in the *Ssrp1* KD by randomly selecting ten intron retention events. The FastStart SYBR Green Master (Roche) was used along with the following primers to amplify via PCR the retained intragenic regions;

Gene name	**Forward primer**	**Reverse primer**
Men1	ATTTCCCAGCAGGCTTCAGG	GGGATGACACGGTTGACAGC
Dvl1	CCTGGGACTACCTCCAGACA	CCTTCATGATGGATCCAATGTA
Map4k2	GCTGCAGTCAGTCCAGGAGG	TCCTGTTGCTTCAGAGTAGCC
Ctsa	GCAATACTCCGGCTACCTCA	TGGGGACTCGATATACAGCA
Pol2ri	CGAAATCGGGAGTGAGTAGC	GGTGGAAGAAGGAACGATCA
Wipf2	TAGAGATGAGCAGCGGAATC	TCGAGAGCTGGGGACTTGCA
Fuz	GACCCAGTGTGTGGACTGTG	GACAAAGGCTGTGCCAGTGG
Rfx5	CACCAGTTGCCCTCTCTGAA	CAATTCTCTTCCTCCCATGC
Fhod1	CACCAGGGAGCAGAGATGAT	CCATCAACATTGGCCTAACC
Tcirg1	AGCGACAGCACTCACTCCTT	CAACACCCCTGCTTCCAGGC

Amplified products were run on a 1.5% Agarose gel and visualized under UV. Band quantification was performed with ImageJ.

Chromatin Immunoprecipitation (ChIP) of FACT subunits. ChIP was performed in ~20 million ES cells, per assay, as described previously(Tessarz et al., 2014) with a few modifications. Briefly, cells were crosslinked with 1% formaldehyde for 20 min followed by quenching for 5 min with the addition of glycine to a final concentration of 0.125 M. After washing with PBS buffer, cells were collected and lysed in Cell Lysis buffer (5 mM Tris pH8.0, 85 mM KCl, 0.5% NP40) with proteinase inhibitors (10 µl/mL Phenylmethylsulfonyl fluoride (PMSF), 1 µl/mL Leupeptin and 1 µl/mL Pepstatin). Pellets were spun for 5 min at 5000 rpm at 4°C. Nuclei were lysed in Nuclei Lysis Buffer (1% SDS, 10 mM EDTA, 50 mM Tris HCl) and samples were sonicated for 12 min. Samples were centrifuged for 20 min at 13,000 rpm at 4°C and the supernatant was diluted in IP buffer (0.01% SDS, 1.1% Triton-X-100, 1.2 mM EDTA, 16.7 mM Tris HCl, 167 mM NaCl) and the appropriate antibody was added and left overnight with rotation at 4°C. Anti-Ssrp1 and anti-Supt16 antibodies were purchased from Biolegends (#609702) and Cell Signalling (#12191) respectively. Two biological replicates were prepared for each FACT subunit, using independent cell cultures and chromatin precipitations. Protein A/G Dynabeads (Invitrogen) were added for 1h and after extensive washed samples were eluted in Elution Buffer (1% SDS, 0.1 M $NaHCO_3$). 20 µL of 5 M NaCl were added and samples were reverse-crosslinked at 65°C for 4h. Following phenol-chloroform extraction and ethanol precipitation, DNA was incubated at 37°C for 4h with RNAse (Sigma).

ChIP-seq library preparation, sequencing, and peak-calling. Approximately 10-20 ng of ChIP material was used for library preparation. End-repair and adaptor ligation was prepared as described previously with a few modifications(Tessarz et al., 2014). Double sided size selections (~200 – 650bp)

were performed using the MagSI-NGS Dynabeads (MagnaMedics, #MD61021) according to the manufacturer's instructions. Purified adapter-ligated ChIP material was run on a high sensitivity DNA chip on a 2200 TapeStation (Agilent) to assess size distribution and adaptor contamination.

Samples were single-end deep-sequenced and reads were aligned to the mm10 genome using Bowtie2 (v 2.2.6) (Langmead & Salzberg, 2012). Peak-calling was performed using PePr (v 1.1) (Zhang, Lin, Johnson, Rozek, & Sartor, 2014) with peaks displaying an FDR < 10^{-5} considered as statistically significant. Peak annotation was performed via the chipenrich(Welch et al., 2014) R package with the following parameters (locusdef = "nearest_gene", method = "broadenrich").

ChIP-seq normalisation and metagene analysis. All the ChIP-seq BAM files were converted to bigwig (10 bp bin) and normalised to x1 sequencing depth using Deeptools (v 2.4)(Ramirez et al., 2016). Blacklisted mm9 co-ordinates were converted to mm10 using the LiftOver tool from UCSC and were further removed from the analysis. Average binding profiles were visualised using R (v 3.3.0). Heatmaps were generated via Deeptools. For the average profiles in Extended Fig 1a,b, RPKM values from Control ES RNA-seq data were divided into four different quantiles and the average profile for each FACT subunit was generated for each quantile. The Pearson's correlation plot in Figure 1a was generated using all unique annotated mm10 RefSeq genes (*n* = 13,348) from UCSC (blacklisted regions were removed). The mESC promoter/enhancer regions identified in Shen *et al*(Shen et al., 2012) were used for the generation of Figure

1b. H3K4me3 ChIP-seq tag densities were split by k-means clustering into three categories in order to remove genes with low H3K4me3 density.

MNase-seq following SSRP1 depletion in mESCs. ES cells were cultured and transfected with shRNA vectors as described above. Biological replicates were obtained from two independent transfection experiments for each shRNA vector. Briefly, ~5 million cells were crosslinked with 1% formaldehyde for 20 min followed by quenching for 5 min with the addition of glycine to a final concentration of 0.125 M. After washing with PBS buffer, cells were collected and lysed in Cell Lysis buffer (5 mM Tris pH8.0, 85 mM KCl, 0.5% NP40) with proteinase inhibitors (10 µl/mL Phenylmethylsulfonyl fluoride (PMSF), 1 µl/mL Leupeptin and 1 µl/mL Pepstatin). Nuclei were gathered by centrifugation (5000 rpm for 2 min) and were treated with 10 Kunitz Units/10^6 cells of micrococcal nuclease (NEB, #M0247S) for 5 min at 37°C. The reaction was stopped with the addition of 60 µl 50 mM EDTA, 25 µl 5 M NaCl, 15 µl 20% NP-40 and incubated on a rotator for 1h at room temperature to release soluble nucleosomes. Samples were centrifuged for 5 min at 10,000 g and supernatant was transferred to a new tube. This centrifugation step is important to obtain highly soluble nucleosomes and remove nucleosome-protein complexes, which can raise bias in subsequent data interpretation(Carone et al., 2014) (See Extended Fig. 8). Samples were reverse-crosslinked by incubating overnight at 65°C with 0.5% SDS and proteinase K. Following phenol-chloroform extraction and ethanol precipitation, DNA was incubated at 37°C for 4h with RNAse (Sigma). All the samples were run in a 2% agarose gel and fragments <200 bp were extracted and purified using the

NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel) according to the manufacturer's instructions.

Purified DNA (500 ng) was used for library preparation as described above. The only difference was the PCR amplification step where we used the same conditions as mentioned in Henikoff *et al* (Henikoff, Belsky, Krassovsky, MacAlpine, & Henikoff, 2011) but with only three amplification cycles. Libraries were verified using a 2200 TapeStation and were paired-end deep-sequenced (~250 million reads per sample).

MNase-seq normalisation and metagene analysis. All the MNase-seq BAM files were converted to bigwig, binned (1 bp), smoothed (20-bp window), and normalised to x1 sequencing depth using Deeptools (v 2.4). Moreover, they were split into two different categories according to fragment length; <80 bp Transcription factor (TF)-sized fragments and 135-170 bp mononucleosome fragments). Average nucleosome occupancy profiles were visualised using R (v 3.3.0). For the Extended Fig 8c, the mm10 annotated exon list for mononucleosomal profiling was obtained from UCSC.

Mass spectrometry sample preparation and analysis. Nuclei were isolated from ~5 million ES cells under hypotonic conditions and samples were incubated overnight at 4°C with an anti-H3K4me3 antibody (Active Motif, #39159) in the presence of low-salt Binding buffer (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 1% NP-40), protease inhibitors, and Protein G Dynabeads (Invitrogen). The following day, after several rounds of bead washing with Binding Buffer, samples were

incubated overnight at 37°C with Tris pH 8.8 and 300 ng Trypsin Gold (Promega). In total, four samples were prepared for each condition (Control, *Ssrp1* KD). Peptides were desalted using StageTips(Rappsilber, Ishihama, & Mann, 2003) and dried. The peptides were resuspended in 0.1% formic acid and analyzed using liquid chromatography - mass spectrometry (LC-MS/MS).

LC-MS/MS analysis. For mass spectrometric analysis, peptides were separated online on a 25 cm 75 µm ID PicoFrit analytical column (New Objective) packed with 1.9 µm ReproSil-Pur media (Dr. Maisch) using an EASY-nLC 1000 (Thermo Fisher Scientific). The column was maintained at 50°C. Buffer A and B were 0.1% formic acid in water and 0.1% formic acid in acetonitrile respectively. Peptides were separated on a segmented gradient from 5% to 25% buffer B for 45 min, from 25% to 35% buffer B for 8 min, and from 35% to 45% buffer B for 4 min, at 200nl / min. Eluting peptides were analyzed on a QExactive HF mass spectrometer (Thermo Fisher Scientific). Peptide precursor mass to charge ratio (m/z) measurements (MS1) were carried out at 60000 resolution in the 300 to 1500 m/z range. The top ten most intense precursors with charge state from 2 to 7 only, were selected for HCD fragmentation using 27% collision energy. The m/z of the peptide fragments (MS2) were measured at 15000 resolution, using an AGC target of 1e6 and 80 ms maximum injection time. Upon fragmentation, precursors were put on an exclusion list for 45 seconds.

LC-MS/MS data analysis. The raw data were analyzed with MaxQuant(Jurgen Cox & Mann, 2008) (v 1.5.2.8) using the integrated Andromeda search engine (Jürgen Cox et al., 2011). Fragmentation spectra were searched against the

canonical and isoform sequences of the mouse reference proteome (proteome ID UP000000589, downloaded August 2015) from UniProt. The database was automatically complemented with sequences of contaminating proteins by MaxQuant. For the data analysis, methionine oxidation and protein N-terminal acetylation were set as variable modifications. The digestion parameters were set to “specific” and “Trypsin/P,” allowing for cleavage after lysine and arginine, also when followed by proline. The minimum number of peptides and razor peptides for protein identification was 1; the minimum number of unique peptides was 0. Protein identification was performed at a peptide spectrum matches and protein false discovery rate of 0.01. The “second peptide” option was on in order to identify co-fragmented peptides. Successful identifications were transferred between the different raw files using the “Match between runs” option, using a match time window of 0.7 min. Label-free quantification (LFQ)(Jurgen Cox, Hein, Luber, & Paron, 2014) was performed using an LFQ minimum ratio count of 2.

Identification of co-enriched proteins. Analysis of the label-free quantification results was done using the Perseus computation platform(Tyanova et al., 2016) (v 1.5.0.0) and R. For the analysis, LFQ intensity values were loaded in Perseus and all identified proteins marked as “Reverse”, “Only identified by site”, and “Potential contaminant” were removed. Upon log2 transformation of the LFQ intensity values, all proteins that contained less than four missing values in one of the groups (control or *Ssrp1* KD) were removed. Missing values in the resulting subset of proteins were imputed with a width of 0.3 and down shift of 1.8. Next, the imputed LFQ intensities were loaded into R where a two side testing for enrichment was performed using limma(Kammers, Cole, Tiengwe, & Ruczinski,

2015; Ritchie et al., 2015). Proteins with an adjusted p-value of less than 0.05 were designated as significantly enriched in the control or knockdown.

NET-seq library preparation. ES cells were cultured and transfected with shRNA vectors as described above. Biological replicates were obtained from two independent transfection experiments for each shRNA vector. NET-seq libraries were prepared as previously described(Mayer & Churchman, 2016) with a few modifications. Briefly, chromatin associated nascent RNA was extracted through cell fractionation in the presence of α-amanitin, protease and RNAase inhibitors. > 90% recovery of ligated RNA and cDNA was achieved from 15 % TBE-Urea (Invitrogen) and 10% TBE-Urea (Invitrogen), respectively, by adding RNA recovery buffer (Zymo Research, R1070-1-10) to the excised gel slices and further incubating at 70°C (1500 rpm) for 15 min. Gel slurry was transferred through a Zymo-Spin IV Column (Zymo Research, C1007-50) and further precipitated for subsequent library preparation steps. cDNA containing the 3' end sequences of a subset of mature and heavily sequenced snRNAs, snoRNAs, and rRNAs, are specifically depleted using biotinylated DNA oligos. Oligo-depleted circularised cDNA was amplified via PCR (5 cycles) and double stranded DNA was run on a 4% low melt agarose gel. The final NET-seq library running at ~150 bp was extracted and further purified using the ZymoClean Gel DNA recovery kit (Zymo Research). Sample purity and concentration was assessed in a 2200 TapeStation and further deep sequenced in a HiSeq 2500 Illumina Platform (~400 million reads per replicate).

NET-seq analysis. All the NET-seq fastq files were processed using custom Python scripts (https://github.com/BradnerLab/netseq) to remove PCR duplicates and reads arising from RT bias. Reads mapping exactly to the last nucleotide of each intron and exon (Splicing intermediates) were further removed from the analysis. The final NET-seq BAM files were converted to bigwig (1 bp bin), separated by strand, and normalised to x1 sequencing depth using Deeptools (v 2.4) with an "–Offset 1" in order to record the position of the 5' end of the sequencing read. NET-seq tags sharing the same or opposite orientation with the TSS were assigned as 'sense' and 'anti-sense' tags, respectively. Average Pol II occupancy profiles were visualised using R (v 3.3.0). In Fig 2g, the Proximal Promoter region was defined as -30 bp and +250 bp around the TSS. For Fig 2f, gene body coverage was retrieved by averaging all regions (FACT-bound and non-FACT-bound) +300 bp downstream of TSS (Transcription Start site) and -200 bp upstream of TES (Transcription End Site). Comparison of the two linear regressions was performed by calculating the z-score via

$$z = \frac{\beta 1 - \beta 2}{\sqrt{s_{\beta 1}^2 + s_{\beta 2}^2}}$$

where β and s_β represent the 'slope' and the 'standard error of the slope', respectively. *P value* was calculated from the respective confidence level yielded by the z score.

Immunofluorescence and confocal microscopy. Early Neuronal Precursor Cells (NPCs) were generated and *Ssrp1* levels were knocked-down as described above. Cells were fixed with 100% Ethanol for 10 min and processed for

immunofluorescence. Permeabilisation and blocking was performed for 1 h at room temperature with 1% BSA and 0.1% NP-40 in PBS. Incubation with primary antibodies was carried at room temperature for 2 hours by using rabbit anti-β3-Tubulin (1:300; Cell Signaling) and mouse anti-MAP2 (1:300; Millipore.). After washing in blocking buffer, the secondary antibodies anti-rabbit and anti-mouse Alexa Fluor 568 (1:1,000; Life Technologies.) were applied for 2 h at room temperature. Slides were extensively washed in PBS and nuclei were counterstained with DAPI before mounting. Fluorescence images were acquired using a laser-scanning confocal microscope (TCS SP5-X; Leica), equipped with a white light laser, a 405-diode UV laser, and a 40× objective lens.

Gene Ontology Analysis. All GO terms were retrieved from the metascape online platform (http://metascape.org/).

Accession numbers and references of publicly available data sets. H3K4me3, H3K27me3, Pol II S5ph, H3K4me1, H3K27Ac, CTCF (ENCODE; https://www.encodeproject.org/); Ep400, Chd1, Chd2, Chd4, Chd6, Chd8, Chd9 (de Dieuleveult et al., 2016) : GSE64825; Set1 (Cxxc1) (Denissov et al., 2014): GSM1258239; p53(Li et al., 2012): GSE26360; Pol II S2ph (Brookes et al., 2016): GSM850470; GRO-seq (Min et al., 2011): GSE27037, Smarcad1 (Xiao et al., 2017) : GSE45338.

8 Literature

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., & Bazett-Jones, D. P. (2010). Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0010531

Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, *17*(8), 487–500. http://doi.org/10.1038/nrg.2016.59

Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., … Ast, G. (2012). Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition. *Cell Reports*, *1*(5), 543–556. http://doi.org/10.1016/j.celrep.2012.03.013

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H. F., John, R. M., … Fisher, A. G. (2006). Chromatin signatures of pluripotent cell lines. *Nat Cell Biol*, *8*(5), 532–538. Retrieved from http://dx.doi.org/10.1038/ncb1403

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M., & Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*, *143*(3), 470–484. http://doi.org/10.1016/j.cell.2010.10.012

Belotserkovskaya, R., Oh, S., Bondarenko, V. a, Orphanides, G., Studitsky, V. M., & Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. *Science (New York, N.Y.)*, *301*(5636), 1090–3. http://doi.org/10.1126/science.1085703

Benayoun, B. A., Pollina, E. A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E. D., … Brunet, A. (2014). H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell*, *158*(3), 673–688. http://doi.org/10.1016/j.cell.2014.06.027

Bibel, M., Richter, J., Lacroix, E., & Barde, Y. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nature Protocols*, *2*(5), 1034–1043. http://doi.org/10.1038/nprot.2007.147

Bieberstein, N. I., Oesterreich, F. C., Straube, K., & Neugebauer, K. M. (2012). First exon length controls active chromatin signatures and transcription. *Cell Reports*, *2*(1), 62–68. http://doi.org/10.1016/j.celrep.2012.05.019

Biswas, D., Takahata, S., Xin, H., Dutta-Biswas, R., Yu, Y., Formosa, T., & Stillman, D. J. (2008). A role for Chd1 and Set2 in negatively regulating DNA replication in Saccharomyces cerevisiae. *Genetics*, *178*(2), 649–659. http://doi.org/10.1534/genetics.107.084202

Biswas, D., Yu, Y., Prall, M., Formosa, T., & Stillman, D. J. (2005). The Yeast FACT Complex Has a Role in Transcriptional Initiation †, *25*(14), 5812–5822. http://doi.org/10.1128/MCB.25.14.5812

Brand, A. H., & Livesey, F. J. (2011). Neural Stem Cell Biology in Vertebrates and Invertebrates: More Alike than Different? *Neuron*, *70*(4), 719–729. http://doi.org/10.1016/j.neuron.2011.05.016

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K. J., Carroll, T., Xie, S. Q., … Pombo, A. (2016). Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs. *Cell Stem Cell*, *10*(2), 157–170. http://doi.org/10.1016/j.stem.2011.12.017

Burckin, T., Nagel, R., Mandel-gutfreund, Y., Shiue, L., Clark, T. A., Chong, J., … Jr, M. A. (2005). Exploring functional relationships between components of the gene expression machinery,

12(2), 175–182. http://doi.org/10.1038/nsmb891

Burgess, R. J., & Zhang, Z. (2013). Histone chaperones in nucleosome assembly and human disease. *Nat Struct Mol Biol*, *20*(1), 14–22. Retrieved from http://dx.doi.org/10.1038/nsmb.2461

Burton, A., & Torres-Padilla, M.-E. (2014). Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nature Reviews. Molecular Cell Biology*, *15*(11), 722–734. http://doi.org/10.1038/nrm3885

Buschbeck, M., & Hake, S. B. (2017). Variants of core histones and their roles in development , stem cells and cancer. *Nature Publishing Group*. http://doi.org/10.1038/nrm.2016.166

Cao, S., Bendall, H., Hicks, G. G., Nashabi, A., Sakano, H., Shinkai, Y., … Ruley, H. E. (2003). The high-mobility-group box protein SSRP1/T160 is essential for cell viability in day 3.5 mouse embryos. *Molecular and Cellular Biology*, *23*(15), 5301–7. http://doi.org/10.1128/MCB.23.15.5301

Carone, B. R., Hung, J. H., Hainer, S. J., Chou, M. Te, Carone, D. M., Weng, Z., … Rando, O. J. (2014). High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Developmental Cell*, *30*(1), 11–22. http://doi.org/10.1016/j.devcel.2014.05.024

Carvalho, S., Raposo, A. C., Martins, F. B., Grosso, A. R., Sridhara, S. C., Rino, J., … De Almeida, S. F. (2013). Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. *Nucleic Acids Research*, *41*(5), 2881–2893. http://doi.org/10.1093/nar/gks1472

Chen, K., Chen, Z., Wu, D., Zhang, L., Lin, X., Su, J., … Li, W. (2015). Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat Genet*, *47*(10), 1149–1157. Retrieved from http://dx.doi.org/10.1038/ng.3385

Cox, J., Hein, M. Y., Luber, C. a, & Paron, I. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular …*, *13*(9), 2513–2526. http://doi.org/10.1074/mcp.M113.031591

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech*, *26*(12), 1367–1372. Retrieved from http://dx.doi.org/10.1038/nbt.1511

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V, & Mann, M. (2011). Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research*, *10*(4), 1794–1805. http://doi.org/10.1021/pr101065j

de Dieuleveult, M., Yen, K., Hmitou, I., Depaux, A., Boussouar, F., Dargham, D. B., … Gérard, M. (2016). Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature*. http://doi.org/10.1038/nature16505

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., … Stewart, A. F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development*, *141*(3), 526 LP-537. Retrieved from http://dev.biologists.org/content/141/3/526.abstract

Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M. M., Heninger, A.-K., Vries, I. de, … Buchholz, F. (2009). A Genome-Scale RNAi Screen for Oct4 Modulators Defines a Role of the Paf1 Complex for Embryonic Stem Cell Identity. *Cell Stem Cell*, *4*(5), 403–415. http://doi.org/10.1016/j.stem.2009.03.009

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. http://doi.org/10.1093/bioinformatics/bts635

Duina, A. A. (2011). Histone Chaperones Spt6 and FACT: Similarities and Differences in Modes of Action at Transcribed Genes. *Genetics Research International*, *2011*, doi:10.4061/2011/625210. http://doi.org/10.4061/2011/625210

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D. J., Dash, C., … Meshorer, E. (2008). Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell*, *2*(5), 437–447. http://doi.org/10.1016/j.stem.2008.03.021

Feng, J., Gan, H., Eaton, M. L., Zhou, H., Li, S., Belsky, J. A., … Li, Q. (2016). Non-coding transcription is a driving force for nucleosome instability in spt16 mutant cells. *Molecular and Cellular Biology*, *36*(13), MCB.00152-16. http://doi.org/10.1128/MCB.00152-16

Formosa, T. (2012). The role of FACT in making and breaking nucleosomes. *Biochimica et Biophysica Acta*, *1819*(3–4), 247–55. http://doi.org/10.1016/j.bbagrm.2011.07.009

Garcia, H., Fleyshman, D., Kolesnikova, K., Safina, A., Commane, M., Paszkiewicz, G., … Gurova, K. (2011). Expression of FACT in mammalian tissues suggests its role in maintaining of undifferentiated state of cells. *Oncotarget*, *2*(10), 783–96. http://doi.org/340 [pii]

Gaspar-Maia, A., Alajem, A., Meshorer, E., & Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. *Nature Reviews Molecular Cell Biology*, *12*(1), 36–47. http://doi.org/10.1038/nrm3036

Gasparian, A. V, Burkhart, C. a, Purmal, A. a, Brodsky, L., Pal, M., Saranadasa, M., … Gurova, K. V. (2011). Curaxins: anticancer compounds that simultaneously suppress NF-κB and activate p53 by targeting FACT. *Science Translational Medicine*, *3*(95), 95ra74. http://doi.org/10.1126/scitranslmed.3002530

Hainer, S. J., Pruneski, J. A., Mitchell, R. D., Monteverde, R. M., & Martens, J. A. (2011). Intergenic transcription causes repression by directing nucleosome assembly. *Genes and Development*, *25*(1), 29–40. http://doi.org/10.1101/gad.1975011

Hautbergue, G. M., Hung, M.-L., Walsh, M. J., Snijders, A. P. L., Chang, C.-T., Jones, R., … Wilson, S. A. (2009). *UIF, a New mRNA Export Adaptor that Works Together with REF/ALY, Requires FACT for Recruitment to mRNA*. *Current Biology* (Vol. 19). Retrieved from http://www.sciencedirect.com/science/article/pii/S0960982209017515

Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., … Ren, B. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, *6*(5), 479–491. http://doi.org/10.1016/j.stem.2010.03.018

Henikoff, J. G., Belsky, J. a., Krassovsky, K., MacAlpine, D. M., & Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. *Proceedings of the National Academy of Sciences*, *108*(45), 18318–18323. http://doi.org/10.1073/pnas.1110731108

Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E. T., … Ladurner, A. G. (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. *Nature*, *499*(7456), 111–4. http://doi.org/10.1038/nature12242

Hsin, J., & Manley, J. L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes and Development*, 2119–2137. http://doi.org/10.1101/gad.200303.112.Transcription

Hynes, R. O. (2012). The evolution of metazoan extracellular matrix. *The Journal of Cell Biology*, *196*(6), 671 LP-679. Retrieved from http://jcb.rupress.org/content/196/6/671.abstract

Iannone, C., Pohl, A., Papasaikas, P., Soronellas, D., Vicent, G. P., Beato, M., & ValcáRcel, J. (2015). Relationship between nucleosome positioning and progesterone-induced alternative splicing in breast cancer cells. *RNA (New York, N.Y.)*, *21*(3), 360–74. http://doi.org/10.1261/rna.048843.114

Ip, J., Schmidt, D., & Pan, Q. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome …*, (416), 390–401. http://doi.org/10.1101/gr.111070.110.390

Jeronimo, C., Watanabe, S., Kaplan, C. D., Peterson, C. L., & Robert, F. (2015). The Histone Chaperones FACT and Spt6 Restrict H2A.Z from Intragenic Locations. *Molecular Cell*, 1–11. http://doi.org/10.1016/j.molcel.2015.03.030

Jonkers, I., Kwak, H., & Lis, J. T. (2014). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *eLife*, *3*(e02407), 1–25. http://doi.org/http://dx.doi.org/10.7554/eLife.02407

Jonkers, I., & Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol*, *16*(3), 167–177. http://doi.org/10.1038/nrm3953

Kadener, S., Cramer, P., Nogués, G., Cazalla, D., de la Mata, M., Fededa, J. P., … Kornblihtt, A. R. (2001). Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. *The EMBO Journal*, *20*(20), 5759 LP-5768. Retrieved from http://emboj.embopress.org/content/20/20/5759.abstract

Kammers, K., Cole, R. N., Tiengwe, C., & Ruczinski, I. (2015). Detecting significant changes in protein abundance. *EuPA Open Proteomics*, *7*, 11–19. http://doi.org/10.1016/j.euprot.2015.02.002

Kasten, M., Szerlong, H., Erdjument-Bromage, H., Tempst, P., Werner, M., & Cairns, B. R. (2004). Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. *The EMBO Journal*, *23*(6), 1348 LP-1359. Retrieved from http://emboj.embopress.org/content/23/6/1348.abstract

Katz, Y., Wang, E. T., Airoldi, E. M., & Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*, *7*(12), 1009–15. http://doi.org/10.1038/nmeth.1528

Kelley, D. E., Stokes, D. G., & Perry, R. P. (1999). CHD1 interacts with SSRP1 and depends on both its chromodomain and its ATPase/helicase-like domain for proper association with chromatin. *Chromosoma*, *108*(1), 10–25. http://doi.org/10.1007/s004120050347

Keogh, M.-C., Kurdistani, S. K., Morris, S. A., Ahn, S. H., Podolny, V., Collins, S. R., … Krogan, N. J. (2005). Cotranscriptional Set2 Methylation of Histone H3 Lysine 36 Recruits a Repressive Rpd3 Complex. *Cell*, *123*(4), 593–605. http://doi.org/10.1016/j.cell.2005.10.025

Koman, I. E., Commane, M., & Paszkiewicz, G. (2012). Targeting FACT Complex Suppresses Mammary Tumorigenesis in Her2 / neu Transgenic Mice, 1025–1035. http://doi.org/10.1158/1940-6207.CAPR-11-0529

Kornblihtt, A. R., Schor, I. E., Allo, M., & Blencowe, B. J. (2009). When chromatin meets splicing. *Nature Structural & Molecular Biology*, *16*(9), 902–903. http://doi.org/10.1038/nsmb0909-902

Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., & Muñoz, M. J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature Reviews. Molecular Cell Biology*, *14*(3), 153–65. http://doi.org/10.1038/nrm3525

Kwak, H., Fuda, N. J., Core, L. J., & Lis, J. T. (2013). Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science*, *339*(6122), 950 LP-953. Retrieved from http://science.sciencemag.org/content/339/6122/950.abstract

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, *9*(4), 357–359. http://doi.org/10.1038/nmeth.1923

Li, M., He, Y., Dubois, W., Wu, X., Shi, J., & Huang, J. (2012). Distinct Regulatory Mechanisms and Functions for p53-Activated and p53-Repressed DNA Damage Response Genes in Embryonic Stem Cells. *Molecular Cell*, *46*(1), 30–42. http://doi.org/10.1016/j.molcel.2012.01.020

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., … Gao, S. (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature*, *537*(7621), 558–562. Retrieved from http://dx.doi.org/10.1038/nature19362

Marom, R., Shur, I., Hager, G. L., & Benayahu, D. (2006). Expression and regulation of CReMM, a chromodomain helicase-DNA-binding (CHD), in marrow stroma derived osteoprogenitors. *Journal of Cellular Physiology*, *207*(3), 628–635. http://doi.org/10.1002/jcp.20611

Mason, P. B., & Struhl, K. (2003). The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Mol Cell Biol*, *23*(22), 8323–8333. http://doi.org/10.1128/MCB.23.22.8323-8333.2003

Mayer, A., & Churchman, L. S. (2016). Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. *Nat Protocols*, *11*(4), 813–833. http://doi.org/10.1038/nprot.2016.047

Mayer, A., Di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., … Churchman, L. S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell*, *161*(3), 541–544. http://doi.org/10.1016/j.cell.2015.03.010

McCullough, L., Rawlins, R., Olsen, A., Xin, H., Stillman, D. J., & Formosa, T. (2011). Insight Into the Mechanism of Nucleosome Reorganization From Histone Mutants That Suppress Defects in the FACT Histone Chaperone. *Genetics*, *188*(4), 835 LP-846. Retrieved from http://www.genetics.org/content/188/4/835.abstract

Min, I. M., Waterfall, J. J., Core, L. J., Munroe, R. J., Schimenti, J., Lis, J. T., … Lis, J. T. (2011). in embryonic stem cells Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes & Development*, *25*(7), 742–754. http://doi.org/10.1101/gad.2005511

Narlikar, G. J., Sundaramoorthy, R., & Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*, *154*(3), 490–503. http://doi.org/10.1016/j.cell.2013.07.011

Neumüller, R. A., Richter, C., Fischer, A., Novatchkova, M., Neumüller, K. G., & Knoblich, J. A. (2011). Genome-wide analysis of self-renewal in Drosophila neural stem cells by transgenic RNAi. *Cell Stem Cell*, *8*(5), 580–593. http://doi.org/10.1016/j.stem.2011.02.022

Nocetti, N., & Whitehouse, I. (2016). Nucleosome repositioning underlies dynamic gene expression. *Genes and Development*, *30*(6), 660–672. http://doi.org/10.1101/gad.274910.115.

Nogues, G., Kadener, S., Cramer, P., Bentley, D., & Kornblihtt, A. R. (2002). Transcriptional activators differ in their abilities to control alternative splicing. *The Journal of Biological Chemistry*, *277*(45), 43110–4. http://doi.org/10.1074/jbc.M208418200

Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., … Proudfoot, N. J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell*, *161*(3), 526–540. http://doi.org/10.1016/j.cell.2015.03.027

Obri, A., Ouararhni, K., Papin, C., Diebold, M.-L., Padmanabhan, K., Marek, M., … Hamiche, A. (2014). ANP32E is a histone chaperone that removes H2A.Z from chromatin. *Nature*, *505*(7485), 648–653. Retrieved from http://dx.doi.org/10.1038/nature12922

Orphanides, G., Wu, W. H., Lane, W. S., Hampsey, M., & Reinberg, D. (1999). The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature*, *400*(6741), 284–8. http://doi.org/10.1038/22350

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., … Thomson, J. A. (2007). Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. *Cell Stem Cell*, *1*(3), 299–312. http://doi.org/10.1016/j.stem.2007.08.003

Park, S.-H., Park, S. H., Kook, M.-C., Kim, E.-Y., Park, S., & Lim, J. H. (2004). Ultrastructure of Human Embryonic Stem Cells and Spontaneous and Retinoic Acid-Induced Differentiating Cells. *Ultrastructural Pathology*, *28*(4), 229–238. http://doi.org/10.1080/01913120490515595

Rakic, P. (2009). Evolution of the neocortex: a perspective from developmental biology. *Nat Rev Neurosci*, *10*(10), 724–735. Retrieved from http://dx.doi.org/10.1038/nrn2719

Ramezani, A., & Hawley, R. G. (2002). Generation of HIV-1-based lentiviral vector particles. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]*, *Chapter 16*, Unit 16.22. http://doi.org/10.1002/0471142727.mb1622s60

Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., … Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(April), 160–165. http://doi.org/10.1093/nar/gkw257

Rappsilber, J., Ishihama, Y., & Mann, M. (2003). Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Analytical Chemistry*, *75*(3), 663–670. http://doi.org/10.1021/ac026117i

Rash, B. G., Lim, H. D., Breunig, J. J., & Vaccarino, F. M. (2011). FGF Signaling Expands Embryonic Cortical Surface Area by Regulating Notch-Dependent Neurogenesis. *The Journal of Neuroscience*, *31*(43), 15604 LP-15617. Retrieved from http://www.jneurosci.org/content/31/43/15604.abstract

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. http://doi.org/10.1093/nar/gkv007

Roberts, G. C., Gooding, C., Mak, H. Y., Proudfoot, N. J., & Smith, C. W. (1998). Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Research*, *26*(24), 5568–5572. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC148035/

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. http://doi.org/10.1093/bioinformatics/btp616

Safina, A., Garcia, H., Commane, M., Guryanova, O., Degan, S., Kolesnikova, K., & Gurova, K. V. (2013). Complex mutual regulation of facilitates chromatin transcription (FACT) subunits on both mRNA and protein levels in human cells. *Cell Cycle*, *12*(15), 2423–2434. http://doi.org/10.4161/cc.25452

Sainsbury, S., Bernecky, C., & Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol*, *16*(3), 129–143. Retrieved from http://dx.doi.org/10.1038/nrm3952

Salomon-Kent, R., Marom, R., John, S., Dundr, M., Schiltz, L. R., Gutierrez, J., … Hager, G. L. (2015). New Face for Chromatin-Related Mesenchymal Modulator: n-CHD9 Localizes to Nucleoli and Interacts With Ribosomal Genes. *Journal of Cellular Physiology*, *230*(9), 2270–80. http://doi.org/10.1002/jcp.24960

Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature*, *489*(7414), 109–113. Retrieved from http://dx.doi.org/10.1038/nature11279

Schor, I. E., Rascovan, N., Pelisch, F., Alló, M., & Kornblihtt, A. R. (2009). Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(11), 4325–30. http://doi.org/10.1073/pnas.0810666106

Scruggs, B. S., Gilchrist, D. A., Nechaev, S., Muse, G. W., Burkholder, A., Fargo, D. C., & Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Molecular Cell*, *58*(6), 1101–1112. http://doi.org/10.1016/j.molcel.2015.04.006

Shao, W., & Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. *Nature Genetics*, (May). http://doi.org/10.1038/ng.3867

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., … Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, *488*(7409), 116–120. http://doi.org/10.1038/nature11243

Sibley, C. R., Blazquez, L., & Ule, J. (2016). Lessons from non-canonical splicing. *Nature Reviews Genetics*, *17*(7), 407–421. http://doi.org/10.1038/nrg.2016.46

Sims, R. J., Millhouse, S., Chen, C. F., Lewis, B. A., Erdjument-Bromage, H., Tempst, P., … Reinberg, D. (2007). Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Molecular Cell*, *28*(4), 665–676. http://doi.org/10.1016/j.molcel.2007.11.010

Talbert, P. B., & Henikoff, S. (2016). Histone variants on the move: substrates for chromatin dynamics. *Nature Reviews. Molecular Cell Biology*. http://doi.org/10.1038/nrm.2016.148

Tan, B. C.-M., Liu, H., Lin, C.-L., & Lee, S.-C. (2010). Functional cooperation between FACT and MCM is coordinated with cell cycle and differential complex formation. *Journal of Biomedical Science*, *17*, 11. http://doi.org/10.1186/1423-0127-17-11

Teif, V. B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.-P., Marth, C., Höfer, T., & Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. *Nature Structural & Molecular Biology*, *19*(11), 1185–1192. http://doi.org/10.1038/nsmb.2419

Tessarz, P., & Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews. Molecular Cell Biology*, *15*(11), 703–708.

http://doi.org/10.1038/nrm3890

Tessarz, P., Santos-Rosa, H., Robson, S. C., Sylvestersen, K. B., Nelson, C. J., Nielsen, M. L., & Kouzarides, T. (2014). Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature*, *505*(7484), 564–8. http://doi.org/10.1038/nature12819

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., … Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*, *489*(7414), 75–82. Retrieved from http://dx.doi.org/10.1038/nature11232

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., & Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol*, *16*(9), 996–1001. http://doi.org/10.1038/nsmb.1658

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., … Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Meth*, *13*(9), 731–740. Retrieved from http://dx.doi.org/10.1038/nmeth.3901

Venkatesh, S., & Workman, J. L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews. Molecular Cell Biology*, *16*(3), 178–189. http://doi.org/10.1038/nrm3941

Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., … Mann, M. (2010). Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell*, *142*(6), 967–980. http://doi.org/10.1016/j.cell.2010.08.020

Vied, C. M., Freudenberg, F., Wang, Y., Raposo, A. a S. F., Feng, D., & Nowakowski, R. S. (2014). A multi-resource data integration approach: identification of candidate genes regulating cell proliferation during neocortical development. *Frontiers in Neuroscience*, *8*(August), 257. http://doi.org/10.3389/fnins.2014.00257

Voong, L. N., Xi, L., Sebeson, A. C., Xiong, B., Wang, J. P., & Wang, X. (2016). Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell*, *167*(6), 1555–1570.e15. http://doi.org/10.1016/j.cell.2016.10.049

Wamstad, J. A., Alexander, J. M., Truty, R. M., Shrikumar, A., Li, F., Eilertson, K. E., … Bruneau, B. G. (2012). Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage. *Cell*, *151*(1), 206–220. http://doi.org/10.1016/j.cell.2012.07.035

Welch, R. P., Lee, C., Imbriano, P. M., Patil, S., Weymouth, T. E., Smith, R. A., … Sartor, M. A. (2014). ChIP-Enrich: Gene set enrichment testing for ChIP-seq data. *Nucleic Acids Research*, *42*(13), 1–13. http://doi.org/10.1093/nar/gku463

Wilson, B. G., & Roberts, C. W. (2011). SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer*, *11*(7), 481–492. http://doi.org/10.1038/nrc3068

Winkler, D. D., & Luger, K. (2011). The Histone Chaperone FACT: Structural Insights and Mechanisms for Nucleosome Reorganization. *Journal of Biological Chemistry*, *286*(21), 18369–18374. http://doi.org/10.1074/jbc.R110.180778

Xiao, S., Lu, J., Sridhar, B., Cao, X., Yu, P., Zhao, T., … Zhong, S. (2017). SMARCAD1 Contributes to the Regulation of Naive Pluripotency by Interacting with Histone Citrullination. *Cell Reports*, *18*(13), 3117–3128. http://doi.org/10.1016/j.celrep.2017.02.070

Zentner, G. E., & Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. *Nature Structural & Molecular Biology*, *20*(3), 259–66.

http://doi.org/10.1038/nsmb.2470

Zhang, Y., Lin, Y. H., Johnson, T. D., Rozek, L. S., & Sartor, M. A. (2014). PePr: A peak-calling prioritization pipeline to identify consistent or differential peaks from replicated ChIP-Seq data. *Bioinformatics*, *30*(18), 2568–2575. http://doi.org/10.1093/bioinformatics/btu372

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K. A., Lozach, J., … Rosenfeld, M. G. (2008). Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation. *Molecular Cell*, *29*(1), 69–80. http://doi.org/10.1016/j.molcel.2007.11.002

9 Acknowledgments

To begin with, I would like to express my immense gratitude to Dr. Peter Tessarz who allowed me to carry out my doctoral thesis in his lab. Without his mentoring and constant intellectual contribution, this project would never have been completed successfully. It's the first time that I work for a PI that I see as a collaborator and not as a boss. Peter, thank you for giving me the freedom to develop and become the independent scientist I am today.

I would also like to thank Ilian Atanassov for carrying out all the Mass-Spectrometry analyses.

Thank you Jorge Boucas, and Sven Templer (MPI-AGE Bioinformatics Core) for all the bioinformatics assistance. I am sorry if I was bugging you more than I should.

To all the members of the Tessarz group; Alex,Julia, Chrysa, Machi, David, Tonantzi, and Nick. Thank you so much guys for the constructive feedback and also for tolerating my grumpiness in dire times.

A HUGE thanks to my parents and sister for constantly believing in me. Max, Franzi, and Sam, thank you for being there for me – your physiological support was the greatest remedy. Και τώρα θα αφιερώσω μια ολόκληρη παράγραφο σε ένα άτομο που πραγματικά με βοήθησε (καλά μην το παίρνεις και πάνω σου) να ολοκληρώσω αυτό το φανταστικό πρότζεκτ. Σε ευχαριστώ για όλη την επαγγελματική και κυρίως την ψυχολογική βοήθεια – θα είχα καταλήξει σε κανένα ψυχιατρείο αν δεν ήσουν εδώ. Ευχαριστώ για όλα τα σούσι, παγωτά και όλες τις βλακείες και παλιμπαιδισμούς (για να μην λές ότι δεν μιλάω σωστά). Δεν θα τα κατάφερνα χωρίς εσένα.

Erklärung zur Dissertation

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt habe, die benutzten Quellen und Hilfsmittel vollstandig angegeben und die Stellen der Arbeit – einschlieslich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universtität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Andreas Beyer betreut worden.

Konstantinos Mylonas

20. April 2018